Supplementary Material: Towards Unbiased and Robust Spatio-Temporal Scene Graph Generation and Anticipation

Anonymous CVPR submission

Paper ID 16783

Appendices

1. Overview

1.1. Motivation

1. **Long-Tailed Distributions**: Real-world visual relationships are dominated by a few frequent *head classes*, while many rare but critical *tail classes* are underrepresented. This imbalance leads to biased models that fail to generalize effectively across all relationship types, compromising nuanced and accurate scene understanding.
2. **Robustness to Distribution Shifts**: Models often struggle with real-world factors such as lighting variations, occlusions, or environmental changes. These distributional shifts degrade performance, limiting the applicability of scene graph models in dynamic and unpredictable environments.

1.2. Motivational Drivers

- **Improving Unbiasedness in Scene Understanding**: A long-tailed distribution skews model learning towards frequent classes, leading to biased predictions. Correctly predicting rare relationships is vital for tasks such as autonomous driving, human-robot interaction, and security surveillance, where underrepresented classes can carry critical contextual information.
- **Enhancing Real-World Applicability**: Distributional shifts are unavoidable in real-world scenarios. Ensuring robustness allows STSG models to remain reliable under practical deployment conditions, bolstering trust and usability.

1.3. Contributions

Thus, concisely, we re-iterate the contributions of the proposed work:

– **Unbiased Learning with Curriculum-Guided Masking**: The proposed **ImparTail** framework leverages curriculum learning and loss masking to prioritize tail classes progressively during training.
– **Introduction of Robustness Metrics**: Two new tasks—*Robust Spatio-Temporal Scene Graph Generation* and *Robust Scene Graph Anticipation*—evaluate model resilience to input corruptions (a step towards analyzing the performance of STSG models under realistic conditions).

2. Extended Related Work

2.1. Structured Visual Representation

Tasks. Learning to represent static visual data like 2D and 3D images as spatial graphs, with objects as nodes and relationships as edges, is called Image Scene Graph Generation (ImgSGG). This field gained momentum with the foundational Visual Genome project [23], advancing 2D ImgSGG research. Building on this foundation, [22] expanded the task to encompass static 3D scene data, including RGB and depth information. Object interactions over time provide richer context when dealing with dynamic visual content like videos. Converting such content into structured Spatio-Temporal Scene Graphs (STSGs), where nodes represent objects and edges capture their temporal relationships, is called Video Scene Graph Generation (VidSGG). The research community has concentrated on improving representation learning through sophisticated object-centric architectures like STTran [5] and RelFormer [34]. These include Open-Vocabulary ImgSGG [4], which expands the range of recognizable objects and relationships. Weakly Supervised ImgSGG [21] to reduce the dependency on extensive labelled data by leveraging weak supervision techniques. Panoptic ImgSGG [47] where panoptic segmentation has been integrated to enhance scene graph representations. Zero-Shot ImgSGG [25, 45] to enable the detection of unseen visual relationships without explicit labels. Shifting gears from identification and generating scene graphs, recently, Peddi et al.[30] introduced the Scene Graph Anticipation (SGA) task to anticipate STSGs for future frames. Alongside these developments, foundation models have advanced various ImgSGG task variants [4, 21, 25, 45, 47].

Unbiased Learning. TEMPURA [28] and FlCoDe[20] address the challenges posed by long-tailed datasets, such as those found in Action Genome [16] and VidVRD [33] and propose methods for unbiased VidSGG. Specifically, FloCoDe [20] mitigates bias by emphasizing temporal consistency and correcting the imbalanced distribution of visual relationships. Similarly, TEMPURA [28] addresses biases in relationship prediction with memory-guided training to generate balanced relationship representations and applies a Gaussian Mixture Model to reduce predictive uncertainty. **Note.** To the best of our knowledge, we are the first to systematically investigate model biases in the SGA task and assess the robustness of both VidSGG and SGA models. In contrast to the above methods, although **ImparTail** shares the goal of training unbiased VidSGG models, it does so without additional architectural components. By modifying the training procedure, **ImparTail** achieves comparable performance and occasionally exceeds the results of these existing methods.

2.2. Learning Paradigms

2.2.1. Long-Tail Learning.

Long-tailed distributions with a few dominant classes (head classes) often overshadow a more significant number of underrepresented ones (tail classes). This class imbalance typically results in models that perform well on head classes but struggle to generalize to tail classes. To mitigate them, the research community has made significant strides in four directions, which include (a) Cost-Sensitive Learning [1, 6, 14, 29, 38, 38, 40, 42, 43], (b) Mixtures-Of-Experts [41] , (c) Resampling Techniques [3, 7, 44], and (d) Specialized Architectures [44, 46].

(a) Cost-Sensitive Learning addresses class imbalance by adjusting the loss function to assign different costs to classes during training. Early approaches involved re-weighting samples inversely proportional to class frequency [14, 38], but this often led to suboptimal performance on real-world data [6]. To improve upon this, advanced methods like label-distribution-aware margin loss with Deferred Re-Weighting (DRW) were proposed [2]. Equalization Loss (EQL) [36] showed that ignoring discouraging gradients for tail classes can prevent adverse effects on model learning. The Class-Balanced (CB) loss [6] re-weights the loss based on the effective number of samples per class, achieving notable performance in single-label classification. Asymmetric Loss (ASL) [1] and Distribution-Balanced (DB) loss [40] focus on balancing positive and negative labels in multi-label classification. Other approaches include transferring knowledge from head to tail classes [29, 38] and designing better training objectives through metric learning [42, 43].

(b) Mixture-of-Experts (MoE) methods tackle class imbalance by distributing samples among specialized expert models. LFME [41] merges multiple experts using self-paced knowledge distillation, while RIDE [37] employs diversity loss and dynamic routing for sample assignment.

(c) Resampling methods adjust the training data distribution by over-sampling tail classes or under-sampling head classes. Techniques like SMOTE [3] create synthetic samples for minority classes, while under-sampling methods reduce samples from majority classes [7]. Approaches, such as concatenating frames from different video clips [44], offer a different way to balance data, particularly spatio-temporal data.

(d) Specialized architectures aim to enhance feature representation and aggregation for tail classes. FrameStack [44] performs frame-level sampling guided by running average precision to improve tail class representation without explicitly differentiating classes at the feature level. The Bilateral-Branch Network (BBN) [46] uses cumulative learning to balance representation learning and classifier discrimination. Kang et al. [19] demonstrated that decoupling representation learning from classifier training prevents head classes from overshadowing tail classes. While Kang et al. [19] argued that resampling might not always be necessary if classifier training is focused correctly, Zhou et al. [46] showed that standard resampling could harm representation learning. Li et al. [26] proposed Gaussian Clouded Logit Adjustment to perturb class logits, adjusting decision boundaries for better generalization across classes.

2.2.2. Curriculum Learning (CL)

is a training methodology that structures training by presenting simpler examples first and progressively introducing more complex ones. This approach aims to enhance learning efficiency by aligning the difficulty of training data with the model's learning capacity at each stage [1][11–13, 24]. Despite its potential benefits, implementing CL presents significant challenges. A primary obstacle is distinguishing between *easy* and *hard* training samples. This differentiation often requires additional mechanisms, such as auxiliary neural networks acting as *teachers* or specialized algorithms. Difficulty measures can be predefined based on certain heuristics [15] or learned automatically during the training process [12, 17, 18, 24, 27, 31, 39]. Alongside difficulty assessment, a scheduling strategy is essential to determine when and how to introduce more challenging data [11]. The *starting small* concept influences our methodology [8], which recommends initiating learning with easier tasks. Unlike conventional Curriculum Learning (CL) methods that introduce data progressively, our approach utilizes the entire training dataset from the start. We adopt label selection to mask the loss function, offering a unique strategy that impacts the learning process while keeping all training examples in play. This approach not only streamlines the implementation of CL but also tackles the difficulties of determining and scheduling the complexity of training data.

[1]It's important to differentiate CL from other dynamic sampling techniques such as self-paced learning [24], boosting [10], hard example mining [35], and active learning [32]. While these methods also adjust the training data based on certain criteria, they typically rely on the model's current performance or hypotheses to select samples rather than following a predefined difficulty progression as in CL.

CVPR
#16783

3. Limitations

1. **Limited Scope of Datasets:** Experiments are primarily conducted on the Action Genome dataset.
   - This is a primary concern of the field as the Action Genome is the only large-scale dataset available as a testbed for the Spatio-Temporal Scene Graph tasks.
2. **Model Robustness to Distribution Shifts:**
   - Although robustness is considered, the specific test corruptions may not cover all real-world scenarios. Instead, our work can be considered a starting point for further developing robust learning techniques.
3. **Bias Mitigation vs. Performance:**
   - In **ImparTail**, balancing unbiased learning with high performance on head classes, although small, might result in a trade-off between performance over head and tail classes. We conjecture that adding an external memory block to our framework can help mitigate this issue.
4. **Limited Evaluation Metrics:**
   - Although metrics such as recall and mean recall provide us insights about the performance of the trained models. These might fail to capture the performance over higher-order spatial and temporal relationships.

4. Approach

4.1. IMPARTAIL

Here, we present the complete algorithm for the proposed unbiased learning framework. Our framework has four key components.

1. (I) **Object Representation Processing Unit (ORPU):** This module extracts object representations for detected objects within video frames using a pre-trained object detector.
2. (II) **Spatio-Temporal Context Processing Unit (STPU):** This unit creates object-centric relationship representations, tailored for two tasks: (i) *observed relationships* for VidSGG and (ii) *anticipated relationships* for SGA.
3. (III) **Relationship Predicate Decoders:** These decoders assign predicate labels to the observed or anticipated relationship representations. **Note:** ORPU, STPU, and the predicate decoders can be adapted from any VidSGG or SGA method following an object-centric framework.
4. (IV) **Curriculum-Guided Masked Loss:** This loss mechanism employs a curriculum-based masking strategy to exclude certain relationship predicate labels during training progressively. Focusing on underrepresented classes helps the model achieve a balanced class distribution.

(IV) Curriculum-Guided Masked Loss. This has two components as explained in the main paper: (a) **Curriculum-Guided Mask Generation** and (b) **Masked Predicate Classification Loss**. First, we provide the complete algorithm for *Mask Generation* and give an overview of the loss function employed. In the subsequent sections, we clearly explain and contrast the loss functions for the original and the proposed IMPARTAIL variants.

4.1.1. Curriculum Guided Mask Generation.

Algorithm 1: Filtered Dataset Construction

Input: Epoch: e, Sampling Ratio: $\mathscr{R}_s$, Dataset Annotations: $\mathcal{D}$, Total predicate labels: N, Total predicates: $\mathcal{P}$, Videos: $\mathscr{V}$
Output: Filtered Dataset: $\mathcal{F}$

*** Determine Target Counts **[can also be a fixed input]** ***
$\mathscr{R}_m = e \times \mathscr{R}_s$ ** Masking Ratio **
$N_{\text{target}} = \text{round}(N \times \mathscr{R}_m)$
*** Curriculum-based sampling probabilities $Prob[rel]$ ***
*** Equally weighted distribution **[can also be learnt]** ***
Set $Prob[rel] = \frac{1}{|\mathcal{P}|}$
Sample target counts $Tar[rel]$ from Multinomial distribution: $Tar[rel] \sim \text{Multinomial}(N_{\text{target}}, Prob[rel])$
*** Randomly sample instances of relationships in the dataset based on the target counts and construct filtered dataset ***
Initialize positions $\mathscr{P}[rel]$ to collect occurrences of rel
foreach v, f *in* $\mathcal{D}$ **do**
  foreach *Relation* rel *in* $\mathcal{D}[v][f]$ **do**
    Append position (v, f, index) to $\mathscr{P}[rel]$
Initialize empty set $\mathcal{K}$
*** These relations are ignored and are omitted from the filtered dataset constructed below ***
foreach *Relation* rel **do**
  Randomly select $Tar[rel]$ positions which should be masked from $\mathscr{P}[rel]$ and add the remaining to $\mathcal{K}$
*** Filter Data ***
Initialize filtered data $\mathcal{F}$
foreach v, f *in* $\mathcal{D}$ **do**
  Initialize filtered frame $\mathcal{F}[v][f]$
  foreach *Relation at position* (v, f, i) **do**
    if $(v, f, i) \in \mathcal{K}$ **then**
      Add relation to $\mathcal{F}[v][f]$

Algorithm 2: Mask Generation

Input: Epoch: e, Dataset Annotations: $\mathcal{D}$, Videos: $\mathscr{V}$, Filtered Dataset: $\mathcal{F}$
Output: Masks: $\mathcal{M}$

Initialize empty mask list $\mathcal{M}$
foreach *Video* v *in* $\mathscr{V}$ **do**
  Initialize video mask $\mathcal{M}_v$
  foreach *Frame* f *in* v **do**
    Initialize frame mask $\mathcal{M}_f$
    foreach *Object* o *in* f **do**
      Initialize object mask $\mathcal{M}_o$
      Original relations $\mathcal{R}_o$ from $\mathcal{D}[v][f]$
      foreach *Relation* rel *in* $\mathcal{R}_o$ **do**
        if $rel \in \mathcal{F}[v][f]$ **then**
          Set mask value $\mathcal{M}_o[rel] = 0$
        else
          Set mask value $\mathcal{M}_o[rel] = 1$
      Add $\mathcal{M}_o$ to $\mathcal{M}_f$
    Add $\mathcal{M}_f$ to $\mathcal{M}_v$
  Add $\mathcal{M}_v$ to $\mathcal{M}^{(e)}$

4.2. Video Scene Graph Generation

4.2.1. IMPARTAIL + STTran

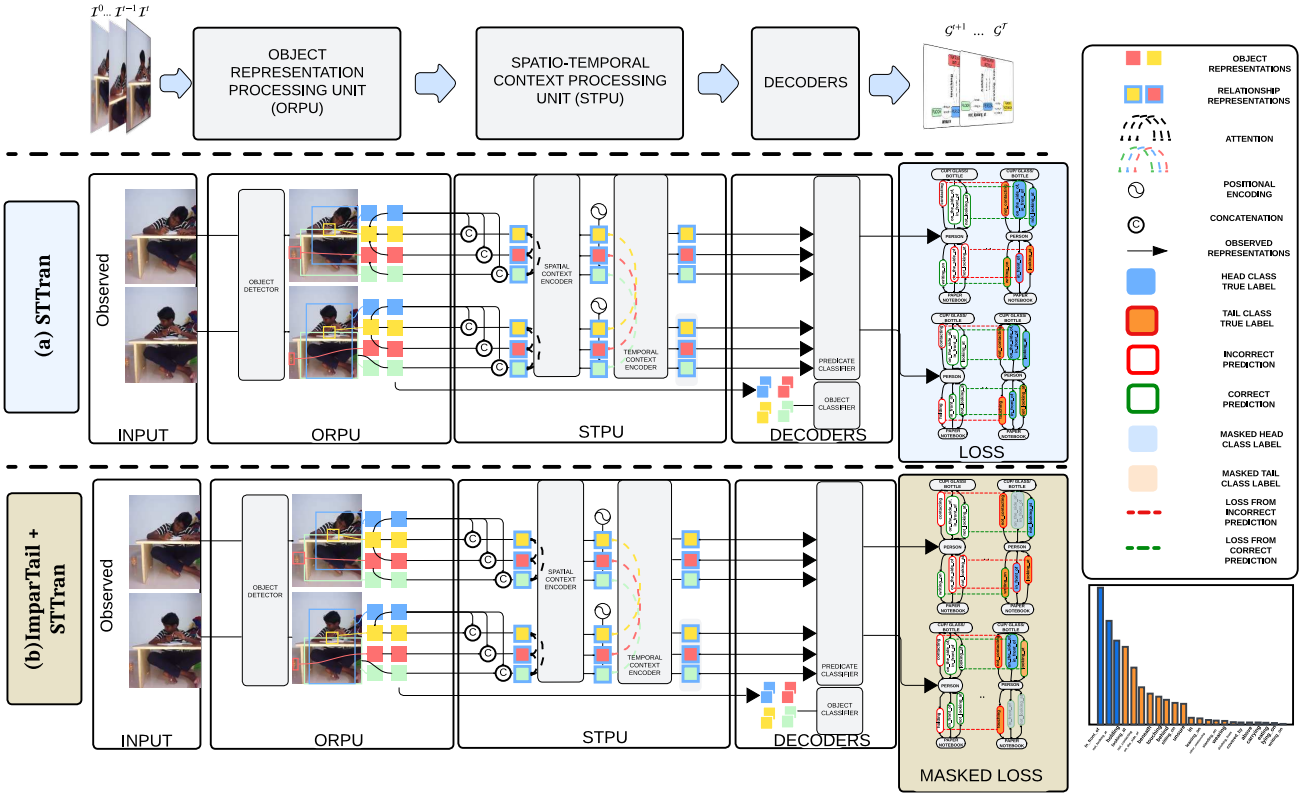

Figure 1. (a) **Architectural Components.** In **STTran**, the Object Representations Processing Unit (ORPU) primarily consists of an object detector and the visual features output by the object detector. Then, the Spatio-Temporal Context Processing Unit (STPU)takes these visual features as input and first constructs relationship representations utilizing the features of interacting objects; then, these relationship representations are fed to a spatial encoder and a transformer encoder. Thus, the spatio-temporal context-aware representations output by the temporal encoder are fed into the predicate classifier for final predictions. Finally, these representations are decoded for predicate classification. (b) **Loss Function.** The primary difference between STTran loss and the proposed IMPARTAIL + STTran loss is illustrated using highlighting the employed losses. We do not mask any predicate label in STTran loss. In contrast, in the proposed IMPARTAIL + STTran loss, we mask the losses corresponding to the *head* classes as generated by following the curriculum-based strategy.

Loss Function STTran

$$\underbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t;}_{(1)} \quad \underbrace{\mathcal{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathcal{L}_{\text{gen}}^t, \mathcal{L}_{\text{gen}}^t = \sum_{ij} \mathcal{L}_{p_{ij}^t}}_{\text{Predicate Classification Loss (2)}}$$

$$\mathcal{L} = \sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathcal{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right)$$

Loss Function IMPARTAIL + STTran

$$\underbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t;}_{(1)} \quad \underbrace{\mathscr{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathscr{L}_{\text{gen}}^t, \mathscr{L}_{\text{gen}}^t = \sum_{ij} \mathbf{m}_{ij}^t * \mathcal{L}_{p_{ij}^t}}_{\textbf{Masked Predicate Classification Loss (2)}} \tag{154}$$

$$\mathcal{L} = \sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathscr{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right) \tag{155}$$

Here, (156)

$$\hat{\mathbf{p}}_{ij}^t = \text{PredClassifier}_{\text{observed}} \left(\mathbf{z}_{ij}^t \right), \forall t \in [1, T] \tag{1}$$ (157)

Predicate Classification Loss ($\mathcal{L}_{\text{gen}}$). focuses on classifying the relationship representations between pairs of objects $\left(o_i^t, o_j^t\right)$ across all frames ($t \in [1, \bar{T}]$) as detailed above. Here, $\mathcal{L}_{p_{ij}^t}$ represents multi-label margin loss and is computed as follows: (158) (159)

$$\mathcal{L}_{p_{ij}^t} = \sum_{u \in \mathcal{P}^+} \sum_{v \in \mathcal{P}^-} \max(0, 1 - \hat{\mathbf{p}}_{ij}^t[v] + \hat{\mathbf{p}}_{ij}^t[u]) \tag{2}$$ (160)

Implementation Details. (161)

– **Training Epochs.** We have capped the number of training epochs for both models where one uses conventional loss and the other uses the proposed IMPARTAIL framework to **5** epochs. (162) (163)

– **Loss Function.** Results reported in the literature for the method STTran were not reproducible using the Multi-Label Margin Loss. However, we noticed we could reach closer numbers (still lower than reported) by employing BCE Loss and training to 10 epochs. (164) (165) (166)

– **Hyperparameters.** We use the same hyperparameter settings described in the paper. (167)

Insight. Our reported mean recall numbers closely match the numbers reported by the SOTA model TEMPURA [28] without any additional architectural changes just by changing how the model is learnt. We also emphasize that although our recall performance was hurt slightly, it is marginally lower than recall values compared to the original model and TEMPURA. (168) (169) (170)

4.2.2. IMPARTAIL + DSGDetr

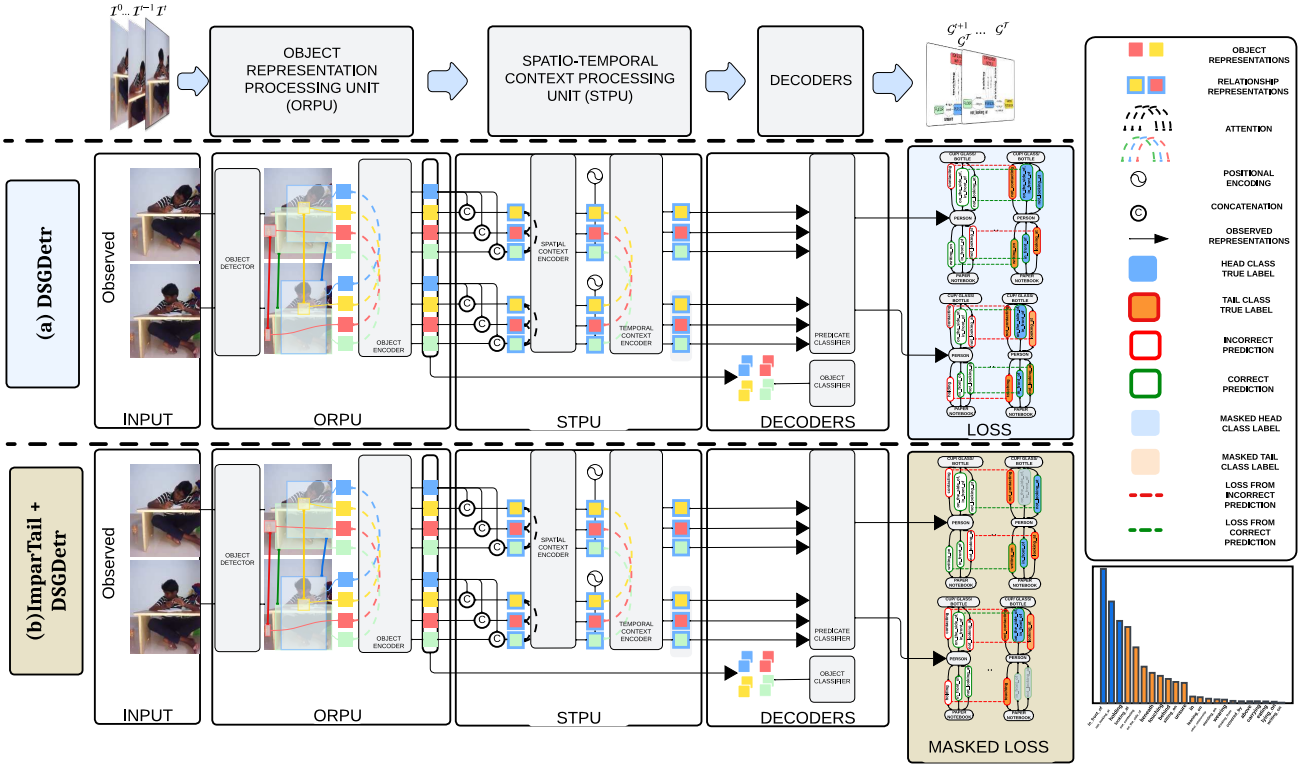

Figure 2. (a) **Architectural Components.** In **DSGDetr**, the Object Representations Processing Unit (ORPU) primarily consists of an object detector, an object tracker and an object encoder. The visual features output by the object detector are used to construct tracklets corresponding to each object, and these representations are further enhanced by passing them through an object encoder. Then, the Spatio-Temporal Context Processing Unit (STPU)takes these visual features as input and first constructs relationship representations utilizing the features of interacting objects; then, these relationship representations are fed to a spatial encoder and a transformer encoder. Thus, the spatio-temporal context-aware representations output by the temporal encoder are fed into the predicate classifier for final predictions. Finally, these representations are decoded for predicate classification. (b) **Loss Function.** The primary difference between DSGDetr loss and the proposed IMPARTAIL + DSGDetr loss is illustrated using highlighting the employed losses. We do not mask any predicate label in DSGDetr loss. In contrast, in the proposed IMPARTAIL + DSGDetr loss, we mask the losses corresponding to the *head* classes as generated by following the curriculum-based strategy.

Loss Function DSGDetr

$$\underbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t;}_{(1)} \quad \underbrace{\mathcal{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathcal{L}_{\text{gen}}^t, \mathcal{L}_{\text{gen}}^t = \sum_{ij} \mathcal{L}_{p_{ij}^t}}_{\text{Predicate Classification Loss (2)}}$$

$$\mathcal{L} = \sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathcal{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right)$$

Loss Function IMPARTAIL + DSGDetr

$$\underbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t;}_{(1)} \quad \underbrace{\mathscr{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathscr{L}_{\text{gen}}^t, \mathscr{L}_{\text{gen}}^t = \sum_{ij} \mathbf{m}_{ij}^t * \mathcal{L}_{p_{ij}^t}}_{\textbf{Masked Predicate Classification Loss (2)}} \tag{174}$$

$$\mathcal{L} = \sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathscr{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right) \tag{175}$$

Here, (176)

$$\hat{\mathbf{p}}_{ij}^t = \text{PredClassifier}_{\text{observed}} \left(\mathbf{z}_{ij}^t \right), \forall t \in [1, T] \tag{3}$$ (177)

Predicate Classification Loss ($\mathcal{L}_{\text{gen}}$). focuses on classifying the relationship representations between pairs of objects $\left(o_i^t, o_j^t\right)$ across all frames ($t \in [1, \bar{T}]$) as detailed above. Here, $\mathcal{L}_{p_{ij}^t}$ represents multi-label margin loss and is computed as follows: (178) (179)

$$\mathcal{L}_{p_{ij}^t} = \sum_{u \in \mathcal{P}^+} \sum_{v \in \mathcal{P}^-} \max(0, 1 - \hat{\mathbf{p}}_{ij}^t[v] + \hat{\mathbf{p}}_{ij}^t[u]) \tag{4}$$ (180)

Implementation Details. (181)

– **Training Epochs.** We have capped the number of training epochs for both models where one uses conventional loss and the other uses the proposed IMPARTAIL framework to **5** epochs. (182) (183)

– **Loss Function.** Results reported in the literature for the method DSGDetr were not reproducible using the Multi-Label Margin Loss. However, we noticed we could reach closer numbers (still lower than reported) by employing BCE Loss and training to 10 epochs. (184) (185) (186)

– **Hyperparameters.** We use the same hyperparameter settings described in the paper. (187)

Insight. Our reported mean recall numbers closely match the numbers reported by the SOTA model TEMPURA [28] without any additional architectural changes just by changing how the model is learnt. We also emphasize that although our recall performance was hurt slightly, it is marginally lower than recall values compared to the original model and TEMPURA. (188) (189) (190)

4.3. Scene Graph Anticipation

4.3.1. IMPARTAIL + STTran++

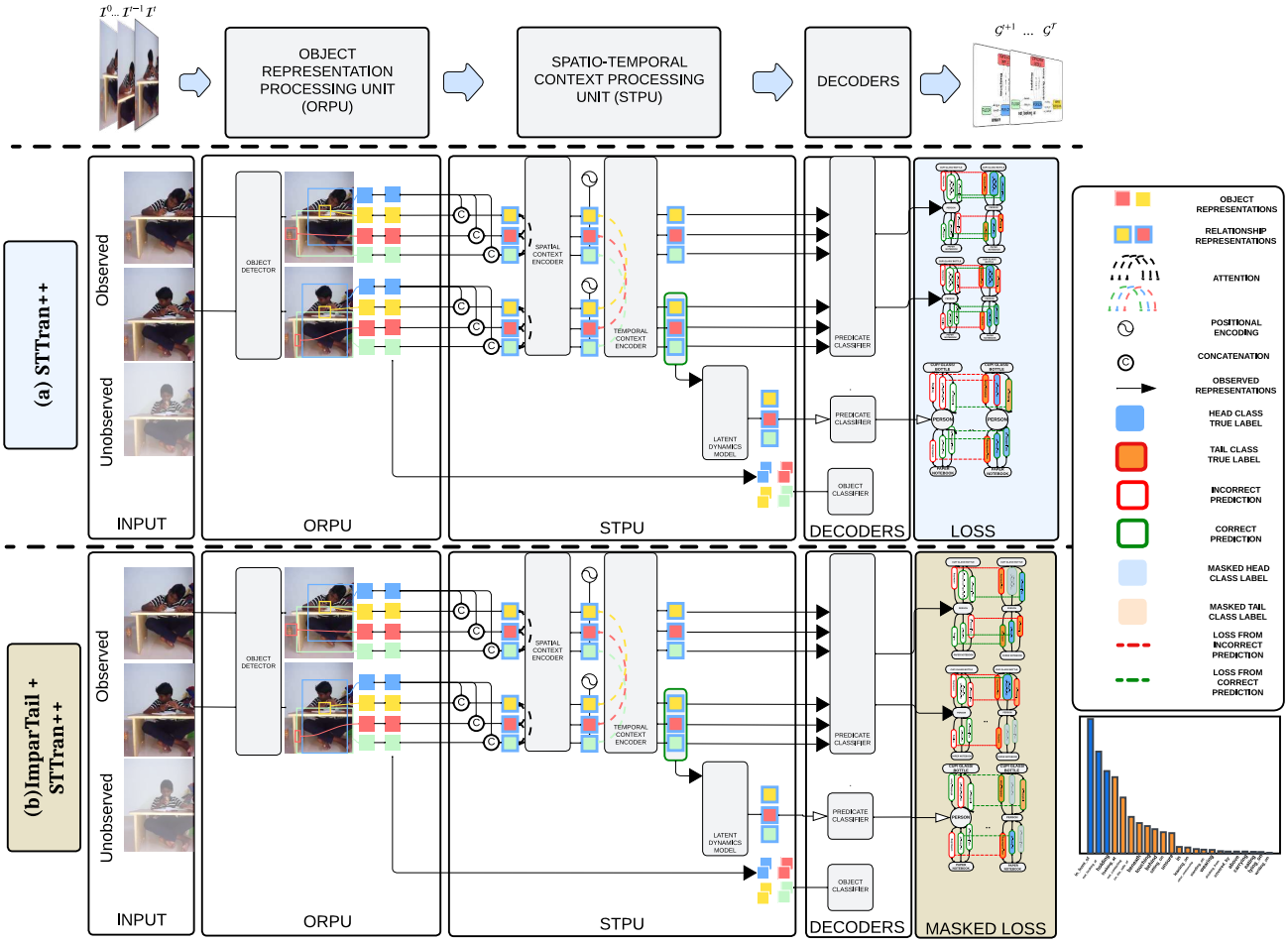

Figure 3. (a) **Architectural Components.** In **STTran++**, the Object Representations Processing Unit (ORPU) primarily consists of an object detector and the visual features output by the object detector. Then, the Spatio-Temporal Context Processing Unit (STPU) takes these visual features as input and first constructs relationship representations utilizing the features of interacting objects; then, these relationship representations are fed to a spatial encoder and a transformer encoder. Thus, the spatio-temporal context-aware representations output by the temporal encoder are fed as input to another transformer encoder to anticipate the future relationship representations corresponding to interacting objects. Thus, relationship representations from the temporal encoder and future relationship representations from the anticipatory transformer encoder are input to two predicate classifiers for final predictions. (b) **Loss Function.** The primary difference between STTran++ loss and the proposed IMPARTAIL + STTran++ loss is illustrated using highlighting the employed losses. We do not mask any predicate label in STTran++ loss. In contrast, in the proposed IMPARTAIL + STTran++ loss, we mask the losses corresponding to the *head* classes output by predicate classification heads corresponding to both observed and anticipated relationship representations.

Loss Function STTran++

$$\mathcal{L}_i = \overbrace{\sum_{t=1}^{\bar{T}} \mathcal{L}_i^t}^{\text{Object Classification Loss (I)}}, \quad \mathcal{L}_i^t = -\sum_{n=1}^{|\mathcal{C}|} y_{i,n}^t \log(\hat{\mathbf{c}}_{i,n}^t) \tag{5}$$

$$\overbrace{\mathcal{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathcal{L}_{\text{gen}}^{t}, \quad \mathcal{L}_{\text{gen}}^{t} = \sum_{ij} \mathcal{L}_{p_{ij}^{t}}}^{\text{Predicate Classification Loss (II)}} \tag{6}$$

$$\underbrace{\mathcal{L}_{\text{ant}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{ant}}^{t}, \quad \mathcal{L}_{\text{ant}}^{t} = \sum_{ij} \mathcal{L}_{p_{ij}^{t}}}_{\text{Predicate Classification Loss (III)}} \tag{7}$$

$$\underbrace{\mathcal{L}_{\text{recon}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{recon}}^{t}, \quad \mathcal{L}_{\text{recon}}^{t} = \frac{1}{N(t) \times N(t)} \sum_{ij}^{(N(t)\times N(t))} \text{L}_{\text{smooth}}(\mathbf{z}_{ij}^{t} - \hat{\mathbf{z}}_{ij}^{t})}_{\text{Reconstruction Loss (IV)}} \tag{8}$$

Thus, the total objective for training the proposed method can be written as:

$$\mathcal{L} = \underbrace{\sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathcal{L}_{\text{gen}}^{t} + \lambda_2 \sum_{i} \mathcal{L}_{i}^{t} \right)}_{\text{Loss Over Observed Representations}} + \underbrace{\sum_{T=3}^{\bar{T}-1} \left(\lambda_3 \mathcal{L}_{\text{ant}}^{(1:T)} + \lambda_4 \mathcal{L}_{\text{recon}}^{(1:T)} \right)}_{\text{Loss Over Anticipated Representations}} \tag{9}$$

Loss Function IMPARTAIL + STTran++

$$\overbrace{\mathcal{L}_{i} = \sum_{t=1}^{\bar{T}} \mathcal{L}_{i}^{t}, \quad \mathcal{L}_{i}^{t} = -\sum_{n=1}^{|\mathcal{C}|} y_{i,n}^{t} \log(\hat{\mathbf{c}}_{i,n}^{t})}^{\text{Object Classification Loss (I)}} \tag{10}$$

$$\overbrace{\mathscr{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathscr{L}_{\text{gen}}^{t}, \quad \mathscr{L}_{\text{gen}}^{t} = \sum_{ij} \mathbf{m}_{ij}^{t} * \mathcal{L}_{p_{ij}^{t}}}^{\text{Predicate Classification Loss (II)}} \tag{11}$$

$$\underbrace{\mathscr{L}_{\text{ant}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathscr{L}_{\text{ant}}^{t}, \quad \mathscr{L}_{\text{ant}}^{t} = \sum_{ij} \mathbf{m}_{ij}^{t} * \mathcal{L}_{p_{ij}^{t}}}_{\text{Predicate Classification Loss (III)}} \tag{12}$$

$$\underbrace{\mathcal{L}_{\text{recon}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{recon}}^{t}, \quad \mathcal{L}_{\text{recon}}^{t} = \frac{1}{N(t) \times N(t)} \sum_{ij}^{(N(t)\times N(t))} \text{L}_{\text{smooth}}(\mathbf{z}_{ij}^{t} - \hat{\mathbf{z}}_{ij}^{t})}_{\text{Reconstruction Loss (IV)}} \tag{13}$$

Thus, the total objective for training the proposed method can be written as:

$$\mathcal{L} = \underbrace{\sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathscr{L}_{\text{gen}}^{t} + \lambda_2 \sum_{i} \mathcal{L}_{i}^{t} \right)}_{\text{Loss Over Observed Representations}} + \underbrace{\sum_{T=3}^{\bar{T}-1} \left(\lambda_3 \mathscr{L}_{\text{ant}}^{(1:T)} + \lambda_4 \mathcal{L}_{\text{recon}}^{(1:T)} \right)}_{\text{Loss Over Anticipated Representations}} \tag{14}$$

Implementation Details.

– **Training Epochs.** We have capped the number of training epochs for both models, where one uses conventional loss, and the other uses the proposed **ImparTail** framework to **5** epochs.

– **Loss Function.** Results reported in the literature for the method STTran++ were reproducible using the Multi-Label Margin Loss. We sometimes achieved higher numbers than those reported in the original paper.

– **Hyperparameters.** We use the same hyperparameter settings described in the paper[30].

4.3.2. ImparTail + DSGDetr++

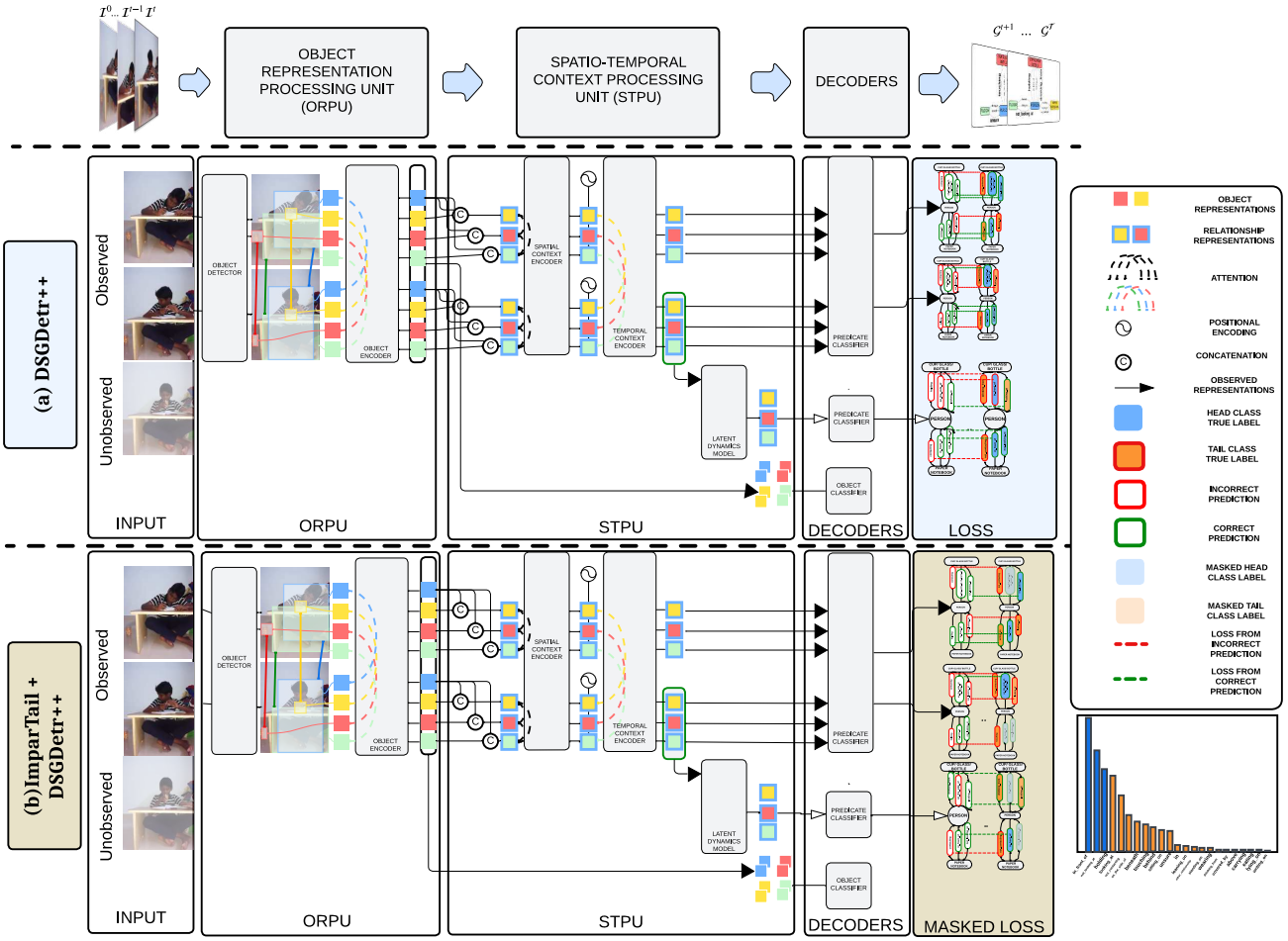

Figure 4. (a) **Architectural Components.** In **DSGDetr++**, the Object Representations Processing Unit (ORPU) primarily consists of an object detector an object tracker and an object encoder. The visual features output by the object detector are used to construct tracklets corresponding to each object, and these representations are further enhanced by passing them through an object encoder. Then, the Spatio-Temporal Context Processing Unit (STPU) takes these visual features as input and first constructs relationship representations utilizing the features of interacting objects; then, these relationship representations are fed to a spatial encoder and a transformer encoder. Thus, the spatio-temporal context-aware representations output by the temporal encoder are fed as input to another transformer encoder to anticipate the future relationship representations corresponding to interacting objects. Thus, relationship representations from the temporal encoder and future relationship representations from the anticipatory transformer encoder are input to two predicate classifiers for final predictions. (b) **Loss Function.** The primary difference between DSGDetr++ loss and the proposed **ImparTail** + DSGDetr++ loss is illustrated using highlighting the employed losses. We do not mask any predicate label in DSGDetr++ loss. In contrast, in the proposed **ImparTail** + DSGDetr++ loss, we mask the losses corresponding to the *head* classes output by predicate classification heads corresponding to both observed and anticipated relationship representations.

Loss Function DSGDetr++

$$\overbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t, \quad \mathcal{L}_i^t = -\sum_{n=1}^{|\mathcal{C}|} y_{i,n}^t \log(\hat{\mathbf{c}}_{i,n}^t)}^{\text{Object Classification Loss (I)}} \tag{15}$$

$$\overbrace{\mathcal{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathcal{L}_{\text{gen}}^t, \quad \mathcal{L}_{\text{gen}}^t = \sum_{ij} \mathcal{L}_{p_{ij}^t}}^{\text{Predicate Classification Loss (II)}} \tag{16}$$

$$\underbrace{\mathcal{L}_{\text{ant}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{ant}}^t, \quad \mathcal{L}_{\text{ant}}^t = \sum_{ij} \mathcal{L}_{p_{ij}^t}}_{\text{Predicate Classification Loss (III)}} \tag{17}$$

$$\underbrace{\mathcal{L}_{\text{recon}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{recon}}^t, \quad \mathcal{L}_{\text{recon}}^t = \frac{1}{N(t) \times N(t)} \sum_{ij}^{(N(t)\times N(t))} \text{L}_{\text{smooth}}(\mathbf{z}_{ij}^t - \hat{\mathbf{z}}_{ij}^t)}_{\text{Reconstruction Loss (IV)}} \tag{18}$$

Thus, the total objective for training the proposed method can be written as:

$$\mathcal{L} = \underbrace{\sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathcal{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right)}_{\text{Loss Over Observed Representations}} + \underbrace{\sum_{T=3}^{\bar{T}-1} \left(\lambda_3 \mathcal{L}_{\text{ant}}^{(1:T)} + \lambda_4 \mathcal{L}_{\text{recon}}^{(1:T)} \right)}_{\text{Loss Over Anticipated Representations}} \tag{19}$$

Loss Function IMPARTAIL + DSGDetr++

$$\overbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t, \quad \mathcal{L}_i^t = -\sum_{n=1}^{|\mathcal{C}|} y_{i,n}^t \log(\hat{\mathbf{c}}_{i,n}^t)}^{\text{Object Classification Loss (I)}} \tag{20}$$

$$\overbrace{\mathscr{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathscr{L}_{\text{gen}}^t, \quad \mathscr{L}_{\text{gen}}^t = \sum_{ij} \mathbf{m}_{ij}^t * \mathcal{L}_{p_{ij}^t}}^{\text{Predicate Classification Loss (II)}} \tag{21}$$

$$\underbrace{\mathscr{L}_{\text{ant}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathscr{L}_{\text{ant}}^t, \quad \mathscr{L}_{\text{ant}}^t = \sum_{ij} \mathbf{m}_{ij}^t * \mathcal{L}_{p_{ij}^t}}_{\text{Predicate Classification Loss (III)}} \tag{22}$$

$$\underbrace{\mathcal{L}_{\text{recon}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{recon}}^t, \quad \mathcal{L}_{\text{recon}}^t = \frac{1}{N(t) \times N(t)} \sum_{ij}^{(N(t)\times N(t))} \text{L}_{\text{smooth}}(\mathbf{z}_{ij}^t - \hat{\mathbf{z}}_{ij}^t)}_{\text{Reconstruction Loss (IV)}} \tag{23}$$

Thus, the total objective for training the proposed method can be written as:

$$\mathcal{L} = \underbrace{\sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathscr{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right)}_{\text{Loss Over Observed Representations}} + \underbrace{\sum_{T=3}^{\bar{T}-1} \left(\lambda_3 \mathscr{L}_{\text{ant}}^{(1:T)} + \lambda_4 \mathcal{L}_{\text{recon}}^{(1:T)} \right)}_{\text{Loss Over Anticipated Representations}} \tag{24}$$

Implementation Details.

– **Training Epochs.** We have capped the number of training epochs for both models, where one uses conventional loss, and the other uses the proposed **ImparTail** framework to **5** epochs.

– **Loss Function.** Results reported in the literature for the method DSGDetr++ were reproducible using the Multi-Label Margin Loss. We sometimes achieved higher numbers than those reported in the original paper.

– **Hyperparameters.** We use the same hyperparameter settings described in the paper[30].

4.3.3. ImparTail + SceneSayer

Loss Function SceneSayer

$$\underbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t, \quad \mathcal{L}_i^t = -\sum_{n=1}^{|\mathcal{C}|} y_{i,n}^t \log(\hat{\mathbf{c}}_{i,n}^t);}_{\text{Object Classification Loss (I)}} \quad \underbrace{\mathcal{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathcal{L}_{\text{gen}}^t, \quad \mathcal{L}_{\text{gen}}^t = \sum_{ij} \mathcal{L}_{p_{ij}^t}}_{\text{Predicate Classification Loss (II)}}$$

$$\mathcal{L}_{\text{ant}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{ant}}^t, \quad \mathcal{L}_{\text{ant}}^t = \sum_{ij} \mathcal{L}_{p_{ij}^t}$$

$$\mathcal{L}_{\text{boxes}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{boxes}}^t, \quad \mathcal{L}_{\text{boxes}}^t = \sum_{k \in \text{boxes}} \text{L}_{\text{smooth}}(b_k^t - \hat{b}_k^t)$$

$$\mathcal{L}_{\text{recon}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{recon}}^t, \quad \mathcal{L}_{\text{recon}}^t = \frac{1}{N(t) \times N(t)} \sum_{ij}^{(N(t) \times N(t))} \text{L}_{\text{smooth}}(\mathbf{z}_{ij}^t - \hat{\mathbf{z}}_{ij}^t)$$

$$\mathcal{L} = \underbrace{\sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathcal{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right)}_{\text{Loss Over Observed Representations}} + \underbrace{\sum_{T=3}^{\bar{T}-1} \left(\lambda_3 \mathcal{L}_{\text{ant}}^{(1:T)} + \lambda_4 \mathcal{L}_{\text{boxes}}^{(1:T)} + \lambda_5 \mathcal{L}_{\text{recon}}^{(1:T)} \right)}_{\text{Loss Over Anticipated Representations}}$$

Loss Function ImparTail + SceneSayer

$$\underbrace{\mathcal{L}_i = \sum_{t=1}^{\bar{T}} \mathcal{L}_i^t, \quad \mathcal{L}_i^t = -\sum_{n=1}^{|\mathcal{C}|} y_{i,n}^t \log(\hat{\mathbf{c}}_{i,n}^t);}_{\text{Object Classification Loss (I)}} \quad \underbrace{\mathscr{L}_{\text{gen}} = \sum_{t=1}^{\bar{T}} \mathscr{L}_{\text{gen}}^t, \quad \mathscr{L}_{\text{gen}}^t = \sum_{ij} \mathbf{m}_{ij}^t * \mathcal{L}_{p_{ij}^t}}_{\text{Predicate Classification Loss (II)}}$$

$$\mathscr{L}_{\text{ant}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathscr{L}_{\text{ant}}^t, \quad \mathscr{L}_{\text{ant}}^t = \sum_{ij} \mathbf{m}_{ij}^t * \mathcal{L}_{p_{ij}^t}$$

$$\mathcal{L}_{\text{boxes}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{boxes}}^t, \quad \mathcal{L}_{\text{boxes}}^t = \sum_{k \in \text{boxes}} \text{L}_{\text{smooth}}(b_k^t - \hat{b}_k^t)$$

$$\mathcal{L}_{\text{recon}}^{(1:T)} = \sum_{t=T+1}^{\min(T+H,\bar{T})} \mathcal{L}_{\text{recon}}^t, \quad \mathcal{L}_{\text{recon}}^t = \frac{1}{N(t) \times N(t)} \sum_{ij}^{(N(t) \times N(t))} \text{L}_{\text{smooth}}(\mathbf{z}_{ij}^t - \hat{\mathbf{z}}_{ij}^t)$$

$$\mathcal{L} = \underbrace{\sum_{t=1}^{\bar{T}} \left(\lambda_1 \mathscr{L}_{\text{gen}}^t + \lambda_2 \sum_i \mathcal{L}_i^t \right)}_{\text{Loss Over Observed Representations}} + \underbrace{\sum_{T=3}^{\bar{T}-1} \left(\lambda_3 \mathscr{L}_{\text{ant}}^{(1:T)} + \lambda_4 \mathcal{L}_{\text{boxes}}^{(1:T)} + \lambda_5 \mathcal{L}_{\text{recon}}^{(1:T)} \right)}_{\text{Loss Over Anticipated Representations}}$$

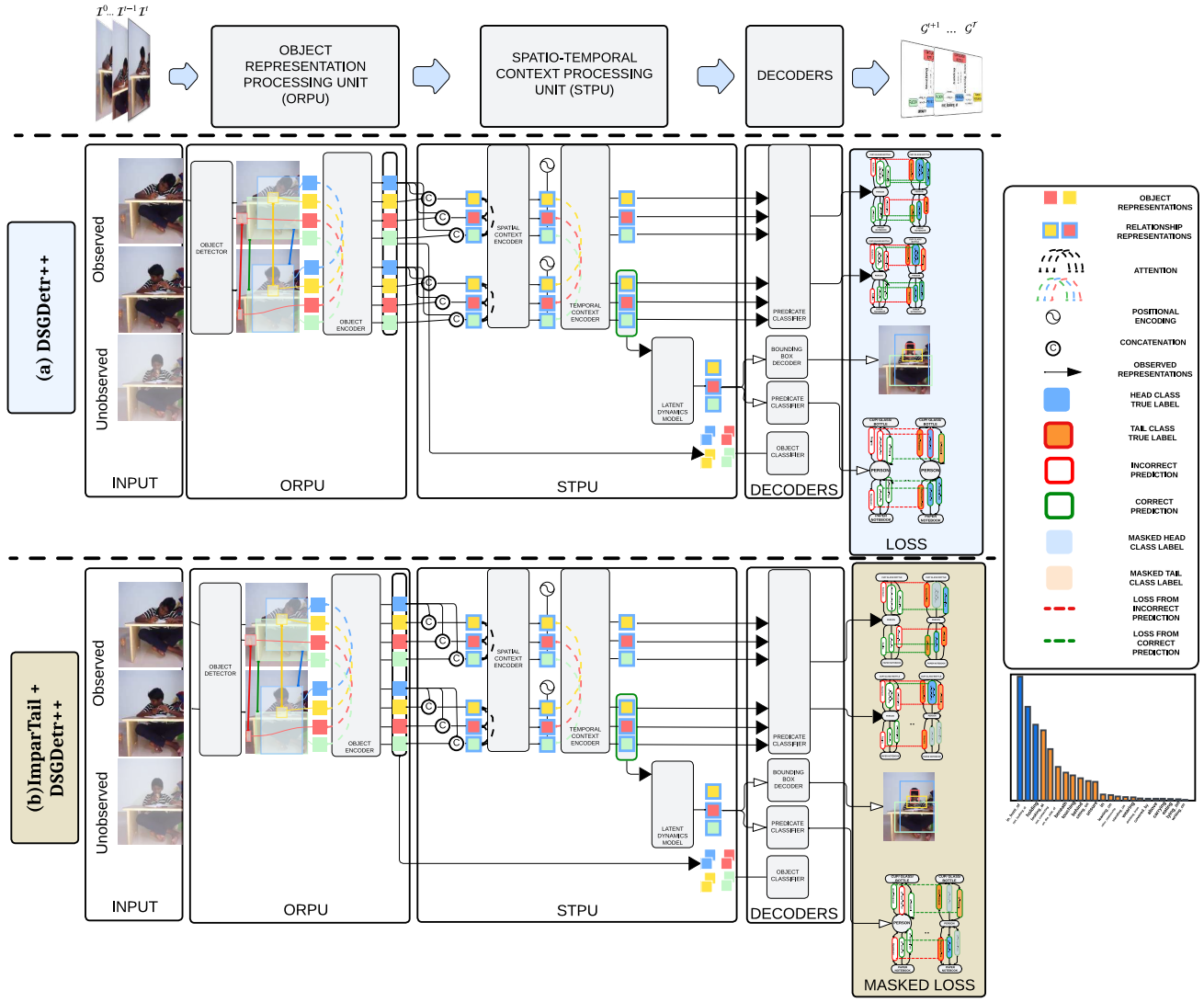

Figure 5. (a) **Architectural Components.** In **SceneSayer**, the Object Representations Processing Unit (ORPU) primarily consists of an object detector an object tracker and an object encoder. The visual features output by the object detector are used to construct tracklets corresponding to each object, and these representations are further enhanced by passing them through an object encoder. Then, the Spatio-Temporal Context Processing Unit (STPU) takes these visual features as input and first constructs relationship representations utilizing the features of interacting objects; then, these relationship representations are fed to a spatial encoder and a transformer encoder. Thus, the spatio-temporal context-aware representations output by the temporal encoder are used as initial values and an Ordinary Differential Equation/ Stochastic Differential Equation is solved to estimate the anticipated future relationship representations corresponding to the interacting objects. Thus, relationship representations from the temporal encoder and future relationship representations from the anticipatory transformer encoder are input to two predicate classifiers for final predictions. (b) **Loss Function.** The primary difference between SceneSayer loss and the proposed IMPARTAIL + SceneSayer loss is illustrated using highlighting the employed losses. We do not mask any predicate label in SceneSayer loss. In contrast, in the proposed IMPARTAIL + SceneSayer loss, we mask the losses corresponding to the *head* classes output by predicate classification heads corresponding to both observed and anticipated relationship representations.

5. Ablation-Overview

5.1. Video Scene Graph Generation

5.1.1. Modes

We evaluate the trained models corresponding to baseline variants STTran , DSGDetr and the proposed method IMPARTAIL + STTran , IMPARTAIL + DSGDetr using three standard modes described in the literature. **(1) Scene Graph Detection (SGDET), (2) Scene Graph Classification (SGCLS) and (3) Predicate Classification (PREDCLS)**.

- **Scene Graph Detection (SGDET):** In this mode, the model is input with frames corresponding to videos. It is tasked to detect objects and predict the relationship predicates between the detected objects.
- **Scene Graph Classification (SGCLS):** In this mode, the model is input with frames corresponding to videos along with bounding boxes of the objects. It is tasked to predict the relationship predicates between the objects.
- **Predicate Classification (PREDCLS):** In this mode, the model is input with frames corresponding to videos along with bounding boxes of the objects and the object labels. It is tasked to predict the relationship predicates between the objects.

5.1.2. Implementation Details.

- **Min Threshold.** As IMPARTAIL proposes a curriculum-guided mask generation strategy, where the number of labels masked in each epoch increases monotonically.
  1. Thus, based on the maximum amount of masking applied, we train three variants of models - {70, 40, 10}.
  2. These models correspond to the following scenarios: (1) **70:** Start from the complete data and reach a {70%, 40%, 10%} masked settings in the last epochs, respectively.
- In section 6, we provide findings corresponding to the proposed training scenarios.

5.2. Scene Graph Anticipation

5.2.1. Modes

We evaluate the trained models corresponding to baseline variants STTran++ , DSGDetr++ , SceneSayerODE , SceneSayerSDE and the proposed methods IMPARTAIL + STTran++ , IMPARTAIL + DSGDetr++ , IMPARTAIL + SceneSayerODE , IMPARTAIL + SceneSayerSDE using three standard modes described in the literature. **(1) Action Genome Scenes (AGS), (2) Partially Grounded Action Genome Scenes (PGAGS) and (3) Grounded Action Genome Scenes (GAGS)**.

- **Action Genome Scenes (AGS):** In this mode, the model receives only the video frames and is tasked to detect objects and infer future relationships between them.
- **Partially Grounded Action Genome Scenes (PGAGS):** In this mode, the model, along with frames, also receives the bounding boxes corresponding to the objects. It is tasked to use this information to infer relationships corresponding to future frames
- **Grounded Action Genome Scenes (GAGS):** In this mode, the model, along with frames, also receives the bounding boxes corresponding to the objects and their labels. It is tasked to use this information to infer future relationships corresponding to interacting objects.

5.2.2. Implementation Details.

- **Min Threshold.** As IMPARTAIL proposes a curriculum-guided mask generation strategy, where the number of labels masked in each epoch increases monotonically.
  1. Thus, based on the maximum amount of masking applied, we train three variants of models - {70, 40, 10}.
  2. These models correspond to the following scenarios: (1) **70:** Start from the complete data and reach a {70%, 40%, 10%} masked settings in the last epochs, respectively.
- In section 6, we provide findings corresponding to the proposed training scenarios.

5.3. Robustness Evaluation.

In section 6, we evaluate the robustness of trained models corresponding to input corruptions and present the results for each mode described above.

6. Ablation Results

6.1. Video Scene Graph Generation

6.1.1. Findings

1. **Table 1** provides a comparative analysis under **NO CONSTRAINT** graph building strategy for different modes and methods for VidSGG, presenting results under various recall metrics (R@10, R@20, R@50, R@100) and mean recall metrics (mR@10, mR@20, mR@50, mR@100).
   (a) We observe that employing the proposed method, the mean recall metrics improved across all modes with only a marginal decrease in recall scores; for example, in the SGDET mode with the STTran method, R@10 slightly decreases from 20.30 to 20.20, conversely, mR@10 increases from 19.30 to 23.50.
   (b) We also observe that the mean recall scores follow a monotonic trend as we increase the masking ratio (avoiding more head classes). We note that the reduction in the recall values is very low.
2. **Table 2, Table 3** provides a comparative analysis under **WITH/SEMI CONSTRAINT** graph building strategy
   (a) We observe that the proposed method improves the mean recall significantly across most setups, though its effect on standard recall metrics is mixed, slightly resulting in a decrease.
3. In Table 2 for STTran under the SGCLS mode, while our method slightly reduced standard recall metrics @100 by 11.6%, there is a notable improvement in mean recall, mR@100 by 36.2%. PREDCLS mode shows less variability in recall changes using our method but substantially increases mean recalls @50 for STTran, jumping from 34.80 to 52.90. While **ImparTail (Ours)** augmentation in constrained and semi-constrained results in decreases in R@50 by 15%, it boosts mR@50 by over 25%.

6.1.2. Results

Table 1. No Constraint Results for VidSGG.

Mode	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
SGCLS	STTran [5]	-	**51.60**	**62.80**	66.30	66.60	38.80	47.10	59.90	66.70
	+ImparTail (Ours)	70	49.80 (-3.49%)	62.40 (-0.64%)	**66.40** (+0.15%)	**66.70** (+0.15%)	43.10 (+11.08%)	53.10 (+12.74%)	61.00 (+1.84%)	65.00 (-2.55%)
	+ImparTail (Ours)	40	49.90 (-3.29%)	62.10 (-1.11%)	66.40 (+0.15%)	66.70 (+0.15%)	45.10 (+16.24%)	55.10 (+16.99%)	64.10 (+7.01%)	66.60 (-0.15%)
	+ImparTail (Ours)	10	48.50 (-6.01%)	61.30 (-2.39%)	66.20 (-0.15%)	66.50 (-0.15%)	**47.40** (+22.16%)	**57.50** (+22.08%)	**66.60** (+11.19%)	**68.10** (+2.10%)
	DSGDetr [9]	-	**55.50**	**68.00**	**72.40**	**72.80**	39.90	49.40	64.60	**72.70**
	+ImparTail (Ours)	70	53.30 (-3.96%)	67.30 (-1.03%)	72.10 (-0.41%)	72.60 (-0.27%)	42.90 (+7.52%)	53.10 (+7.49%)	66.10 (+2.32%)	72.20 (-0.69%)
	+ImparTail (Ours)	40	50.90 (-8.29%)	66.10 (-2.79%)	72.10 (-0.41%)	72.60 (-0.27%)	45.00 (+12.78%)	55.50 (+12.35%)	67.20 (+4.02%)	71.40 (-1.79%)
	+ImparTail (Ours)	10	50.20 (-9.55%)	65.30 (-3.97%)	71.90 (-0.69%)	72.70 (-0.14%)	**48.80** (+22.31%)	**59.60** (+20.65%)	**70.10** (+8.51%)	72.20 (-0.69%)
SGDET	STTran [5]	-	**20.30**	**31.10**	**45.90**	48.40	19.30	26.90	35.60	39.70
	+ImparTail (Ours)	70	19.80 (-2.46%)	30.20 (-2.89%)	45.80 (-0.22%)	**48.60** (+0.41%)	20.80 (+7.77%)	29.50 (+9.67%)	38.70 (+8.71%)	42.00 (+5.79%)
	+ImparTail (Ours)	40	20.20 (-0.49%)	30.80 (-0.96%)	45.30 (-1.31%)	48.20 (-0.41%)	22.60 (+17.10%)	31.10 (+15.61%)	39.10 (+9.83%)	42.10 (+6.05%)
	+ImparTail (Ours)	10	20.00 (-1.48%)	30.10 (-3.22%)	45.10 (-1.74%)	48.50 (+0.21%)	**23.50** (+21.76%)	**33.60** (+24.91%)	**43.80** (+23.03%)	**47.00** (+18.39%)
	DSGDetr [9]	-	**29.80**	**39.00**	**46.40**	48.30	23.30	29.80	36.00	39.70
	+ImparTail (Ours)	70	28.60 (-4.03%)	38.00 (-2.56%)	46.40	**49.00** (+1.45%)	25.00 (+7.30%)	32.00 (+7.38%)	39.30 (+9.17%)	42.50 (+7.05%)
	+ImparTail (Ours)	40	25.90 (-13.09%)	36.10 (-7.44%)	45.60 (-1.72%)	48.70 (+0.83%)	24.70 (+6.01%)	32.20 (+8.05%)	39.60 (+10.00%)	43.20 (+8.82%)
	+ImparTail (Ours)	10	26.50 (-11.07%)	36.10 (-7.44%)	45.20 (-2.59%)	48.40 (+0.21%)	**27.50** (+18.03%)	**35.20** (+18.12%)	**43.30** (+20.28%)	**46.60** (+17.38%)
PREDCLS	STTran [5]	-	**73.20**	**92.70**	99.20	**99.90**	45.70	63.40	80.50	95.60
	+ImparTail (Ours)	70	70.20 (-4.10%)	91.40 (-1.40%)	**99.30** (+0.10%)	99.90	55.00 (+20.35%)	71.80 (+13.25%)	86.70 (+7.70%)	97.00 (+1.46%)
	+ImparTail (Ours)	40	67.50 (-7.79%)	89.70 (-3.24%)	99.20	99.90	54.80 (+19.91%)	72.10 (+13.72%)	86.70 (+7.70%)	97.20 (+1.67%)
	+ImparTail (Ours)	10	67.50 (-7.79%)	88.80 (-4.21%)	99.00 (-0.20%)	99.90	**65.50** (+43.33%)	**82.00** (+29.34%)	**93.00** (+15.53%)	**99.60** (+4.18%)
	DSGDetr [9]	-	**72.80**	**92.40**	**99.20**	**99.90**	45.60	64.40	80.50	94.70
	+ImparTail (Ours)	70	67.70 (-7.01%)	89.60 (-3.03%)	99.20	99.90	56.00 (+22.81%)	72.60 (+12.73%)	85.90 (+6.71%)	97.30 (+2.75%)
	+ImparTail (Ours)	40	68.00 (-6.59%)	90.10 (-2.49%)	99.20	99.90	54.50 (+19.52%)	71.80 (+11.49%)	86.40 (+7.33%)	97.30 (+2.75%)
	+ImparTail (Ours)	10	65.80 (-9.62%)	87.70 (-5.09%)	98.90 (-0.30%)	99.90	**59.40** (+30.26%)	**76.20** (+18.32%)	**89.80** (+11.55%)	**98.10** (+3.59%)

Table 2. With Constraint Results for VidSGG.

Mode	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
SGCLS	STTran [5]	-	**44.90**	**46.50**	**46.50**	**46.50**	25.00	27.50	27.60	27.60
	+ImpartTail (Ours)	70	42.00 (-6.46%)	43.30 (-6.88%)	43.30 (-6.88%)	43.30 (-6.88%)	25.90 (+3.60%)	28.70 (+4.36%)	28.80 (+4.35%)	28.80 (+4.35%)
	+ImpartTail (Ours)	40	42.40 (-5.57%)	43.70 (-6.02%)	43.80 (-5.81%)	43.80 (-5.81%)	27.80 (+11.20%)	30.60 (+11.27%)	30.70 (+11.23%)	30.70 (+11.23%)
	+ImpartTail (Ours)	10	39.90 (-11.14%)	41.10 (-11.61%)	41.10 (-11.61%)	41.10 (-11.61%)	**32.30** (+29.20%)	**36.20** (+31.64%)	**36.20** (+31.16%)	**36.20** (+31.16%)
	DSGDetr [9]	-	**47.80**	**49.30**	**49.40**	**49.40**	25.60	28.10	28.10	28.10
	+ImpartTail (Ours)	70	46.00 (-3.77%)	47.40 (-3.85%)	47.40 (-4.05%)	47.40 (-4.05%)	27.10 (+5.86%)	30.20 (+7.47%)	30.30 (+7.83%)	30.30 (+7.83%)
	+ImpartTail (Ours)	40	41.20 (-13.81%)	42.40 (-14.00%)	42.40 (-14.17%)	42.40 (-14.17%)	27.90 (+8.98%)	30.80 (+9.61%)	30.80 (+9.61%)	30.80 (+9.61%)
	+ImpartTail (Ours)	10	40.50 (-15.27%)	42.00 (-14.81%)	42.00 (-14.98%)	42.00 (-14.98%)	**32.20** (+25.78%)	**36.00** (+28.11%)	**36.00** (+28.11%)	**36.00** (+28.11%)
SGDET	STTran [5]	-	**19.00**	**29.40**	**32.10**	**32.10**	8.00	16.60	19.30	19.30
	+ImpartTail (Ours)	70	17.90 (-5.79%)	27.80 (-5.44%)	30.60 (-4.67%)	30.60 (-4.67%)	8.20 (+2.50%)	17.50 (+5.42%)	20.60 (+6.74%)	20.60 (+6.74%)
	+ImpartTail (Ours)	40	17.50 (-7.89%)	27.50 (-6.46%)	30.30 (-5.61%)	30.30 (-5.61%)	8.80 (+10.00%)	19.20 (+15.66%)	22.60 (+17.10%)	22.60 (+17.10%)
	+ImpartTail (Ours)	10	16.00 (-15.79%)	25.60 (-12.93%)	28.40 (-11.53%)	28.40 (-11.53%)	**9.40** (+17.50%)	**21.50** (+29.52%)	**25.90** (+34.20%)	**25.90** (+34.20%)
	DSGDetr [9]	-	**17.10**	**28.80**	**33.90**	**33.90**	6.70	14.70	19.10	19.10
	+ImpartTail (Ours)	70	16.30 (-4.68%)	27.50 (-4.51%)	32.50 (-4.13%)	32.60 (-3.83%)	7.40 (+10.45%)	17.60 (+19.73%)	23.20 (+21.47%)	23.20 (+21.47%)
	+ImpartTail (Ours)	40	14.10 (-17.54%)	23.40 (-18.75%)	27.40 (-19.17%)	27.50 (-18.88%)	7.30 (+8.96%)	16.80 (+14.29%)	22.40 (+17.28%)	22.40 (+17.28%)
	+ImpartTail (Ours)	10	15.40 (-9.94%)	25.70 (-10.76%)	30.10 (-11.21%)	30.10 (-11.21%)	**7.50** (+11.94%)	**17.80** (+21.09%)	**23.70** (+24.08%)	**23.80** (+24.61%)
PREDCLS	STTran [5]	-	**66.40**	**69.90**	**69.90**	**69.90**	30.50	34.70	34.80	34.80
	+ImpartTail (Ours)	70	61.50 (-7.38%)	64.80 (-7.30%)	64.80 (-7.30%)	64.80 (-7.30%)	34.30 (+12.46%)	39.70 (+14.41%)	39.80 (+14.37%)	39.80 (+14.37%)
	+ImpartTail (Ours)	40	57.20 (-13.86%)	60.20 (-13.88%)	60.30 (-13.73%)	60.30 (-13.73%)	37.40 (+22.62%)	43.60 (+25.65%)	43.80 (+25.86%)	43.80 (+25.86%)
	+ImpartTail (Ours)	10	57.70 (-13.10%)	60.80 (-13.02%)	60.80 (-13.02%)	60.80 (-13.02%)	**44.00** (+44.26%)	**52.70** (+51.87%)	**52.90** (+52.01%)	**52.90** (+52.01%)
	DSGDetr [9]	-	**66.50**	**70.00**	**70.00**	**70.00**	31.50	36.10	36.20	36.20
	+ImpartTail (Ours)	70	58.30 (-12.33%)	61.50 (-12.14%)	61.50 (-12.14%)	61.50 (-12.14%)	38.20 (+21.27%)	45.00 (+24.65%)	45.10 (+24.59%)	45.10 (+24.59%)
	+ImpartTail (Ours)	40	58.00 (-12.78%)	61.10 (-12.71%)	61.10 (-12.71%)	61.10 (-12.71%)	37.30 (+18.41%)	43.40 (+20.22%)	43.50 (+20.17%)	43.50 (+20.17%)
	+ImpartTail (Ours)	10	55.50 (-16.54%)	58.30 (-16.71%)	58.30 (-16.71%)	58.30 (-16.71%)	**41.00** (+30.16%)	**48.10** (+33.24%)	**48.20** (+33.15%)	**48.20** (+33.15%)

Table 3. Semi Constraint Results for VidSGG.

Mode	Method	$\mathcal{S}$	SEMI CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
SGCLS	STTran [5]	-	**49.90**	**55.80**	**56.20**	**56.20**	29.50	39.90	40.90	40.90
	+ImpartTail (Ours)	70	49.00 (-1.80%)	55.60 (-0.36%)	56.20	56.20	32.50 (+10.17%)	45.80 (+14.79%)	47.60 (+16.38%)	47.60 (+16.38%)
	+ImpartTail (Ours)	40	48.60 (-2.61%)	54.80 (-1.79%)	55.20 (-1.78%)	55.20 (-1.78%)	34.30 (+16.27%)	48.40 (+21.30%)	50.00 (+22.25%)	50.00 (+22.25%)
	+ImpartTail (Ours)	10	46.40 (-7.01%)	52.40 (-6.09%)	52.80 (-6.05%)	52.80 (-6.05%)	**36.20** (+22.71%)	**50.50** (+26.57%)	**52.20** (+27.63%)	**52.20** (+27.63%)
	DSGDetr [9]	-	**53.90**	**60.40**	**60.70**	**60.70**	30.10	40.60	41.60	41.60
	+ImpartTail (Ours)	70	52.30 (-2.97%)	59.60 (-1.32%)	60.30 (-0.66%)	60.30 (-0.66%)	32.50 (+7.97%)	45.20 (+11.33%)	47.20 (+13.46%)	47.20 (+13.46%)
	+ImpartTail (Ours)	40	50.50 (-6.31%)	58.50 (-3.15%)	59.50 (-1.98%)	59.50 (-1.98%)	33.90 (+12.62%)	49.00 (+20.69%)	51.70 (+24.28%)	51.70 (+24.28%)
	+ImpartTail (Ours)	10	46.80 (-13.17%)	53.80 (-10.93%)	54.40 (-10.38%)	54.40 (-10.38%)	**36.80** (+22.26%)	**52.40** (+29.06%)	**54.90** (+31.97%)	**54.90** (+31.97%)
SGDET	STTran [5]	-	**18.60**	**31.00**	**41.20**	41.50	7.70	18.20	30.40	30.80
	+ImpartTail (Ours)	70	18.00 (-3.23%)	30.10 (-2.90%)	41.10 (-0.24%)	**41.70** (+0.48%)	7.90 (+2.60%)	19.00 (+4.40%)	33.90 (+11.51%)	34.60 (+12.34%)
	+ImpartTail (Ours)	40	17.70 (-4.84%)	29.70 (-4.19%)	39.10 (-5.10%)	39.40 (-5.06%)	**8.70** (+12.99%)	21.30 (+17.03%)	34.80 (+14.47%)	35.10 (+13.96%)
	+ImpartTail (Ours)	10	16.40 (-11.83%)	28.20 (-9.03%)	37.90 (-8.01%)	38.20 (-7.95%)	8.60 (+11.69%)	**21.80** (+19.78%)	**38.30** (+25.99%)	**38.80** (+25.97%)
	DSGDetr [9]	-	**16.40**	**28.70**	**40.70**	**41.50**	6.50	16.00	30.40	31.50
	+ImpartTail (Ours)	70	15.80 (-3.66%)	27.90 (-2.79%)	40.20 (-1.23%)	41.20 (-0.72%)	6.90 (+6.15%)	17.30 (+8.12%)	34.20 (+12.50%)	35.60 (+13.02%)
	+ImpartTail (Ours)	40	14.10 (-14.02%)	25.20 (-12.20%)	37.30 (-8.35%)	38.70 (-6.75%)	6.90 (+6.15%)	16.70 (+4.38%)	33.40 (+9.87%)	35.20 (+11.75%)
	+ImpartTail (Ours)	10	15.20 (-7.32%)	26.80 (-6.62%)	37.90 (-6.88%)	39.00 (-6.02%)	**7.30** (+12.31%)	**18.40** (+15.00%)	**36.60** (+20.39%)	**38.40** (+21.90%)
PREDCLS	STTran [5]	-	**71.80**	**82.50**	**83.30**	**83.30**	36.60	51.80	53.80	53.80
	+ImpartTail (Ours)	70	69.60 (-3.06%)	81.30 (-1.45%)	82.60 (-0.84%)	82.60 (-0.84%)	41.60 (+13.66%)	61.90 (+19.50%)	65.60 (+21.93%)	65.60 (+21.93%)
	+ImpartTail (Ours)	40	66.40 (-7.52%)	77.90 (-5.58%)	79.30 (-4.80%)	79.30 (-4.80%)	42.10 (+15.03%)	62.40 (+20.46%)	66.20 (+23.05%)	66.20 (+23.05%)
	+ImpartTail (Ours)	10	63.80 (-11.14%)	74.20 (-10.06%)	75.20 (-9.72%)	75.20 (-9.72%)	**47.70** (+30.33%)	**69.70** (+34.56%)	**73.40** (+36.43%)	**73.40** (+36.43%)
	DSGDetr [9]	-	**71.30**	**82.50**	**83.50**	**83.50**	36.50	52.50	55.20	55.20
	+ImpartTail (Ours)	70	66.80 (-6.31%)	78.90 (-4.36%)	80.40 (-3.71%)	80.40 (-3.71%)	41.60 (+13.97%)	63.00 (+20.00%)	67.70 (+22.64%)	67.70 (+22.64%)
	+ImpartTail (Ours)	40	66.10 (-7.29%)	77.30 (-6.30%)	78.70 (-5.75%)	78.70 (-5.75%)	42.30 (+15.89%)	61.70 (+17.52%)	65.30 (+18.30%)	65.30 (+18.30%)
	+ImpartTail (Ours)	10	63.50 (-10.94%)	75.50 (-8.48%)	77.10 (-7.66%)	77.10 (-7.66%)	**43.90** (+20.27%)	**65.40** (+24.57%)	**69.80** (+26.45%)	**69.80** (+26.45%)

6.2. Scene Graph Anticipation

6.2.1. Findings

Here, $\mathcal{S}$ - represents the amount of things included in the masked dataset. To be more precise, $\mathcal{S} = 10\%$ means that only 10% of the labels are included in training the model for the current epoch and 90% of the labels are masked, thus voiding their contribution to the loss. So $\mathcal{S} = 70\%$ has more labels contributing to the training loss and $\mathcal{S} = 10\%$ less number of labels contributing to the training loss.

1. Table 4, Table 5, Table 6, Table 7, compare proposed method's performance across various base methods (STTran++, DSGDet++, SceneSayerODE, and SceneSayerSDE) at different $\mathcal{F}$ values (0.3, 0.5, 0.7, and 0.9) for Scene Graph Generation (SGA) task, under the GAGS-No Constraint setting.
   (a) In Table 4, SceneSayerODE shows the most consistent gain in lower recall metrics (R@10 and mR@10) when **ImparTail** is included. For $\mathcal{S}$=70, improvements with **ImparTail** are substantial, especially for STTran++ (e.g., +13.74% for R@10). Lower values of $\mathcal{S}$ (e.g., $\mathcal{S}$=10) tend to result in less significant improvements. Metrics like mR@50 and mR@100 remain stable or show slight improvements, emphasizing **ImparTail**'s balanced handling of long-tail distributions.
   (b) In Table 5, Table 6, Table 7, **ImparTail** continues to show consistent improvements, especially for mR metrics, with substantial gains seen in SceneSayerODE and SceneSayerSDE for mR@10 and mR@20. The relative improvement in metrics is more pronounced compared to Table 4, suggesting that **ImparTail** is more impactful as the $\mathcal{F}$ value increases.
2. Table 8, Table 9, Table 10, Table 11, evaluates **ImparTail** under the Partially Grounded Action Genome Scenes (PGAGS) - No Constraint setting.
   (a) As $\mathcal{F}$ increases from 0.3 to 0.9, the improvements in mR metrics, particularly for mR@10 and mR@20, become more pronounced.
   (b) **ImparTail** improves both R metrics (favouring head classes) and mR metrics (favouring tail classes) as $\mathcal{F}$ increases. For instance, significant gains in mR@10 and mR@20 consistently align with moderate or stable improvements in R@10 and R@20 across all models and configurations.
   (c) All baseline methods (STTran++, DSGDet++, SceneSayerODE, and SceneSayerSDE) benefit from the inclusion of **ImparTail (Ours)**, though the degree of improvement varies. The results also highlight **ImparTail**'s compatibility with high $\mathcal{S}$, with significant gains observed at high $\mathcal{F}$ values (e.g., +49.76% for mR@10 at F=0.7).
3. Table 12, Table 13, Table 14, Table 15, present the performance evaluation of **ImparTail**'s under the Action Genome Scenes (AGS) in No Constraint graph building strategy, for Scene Graph Generation (SGA).
   (a) At lower $\mathcal{F}$ values (e.g., $\mathcal{F}$=0.3), the improvements in mR metrics are moderate. At higher $\mathcal{F}$ values (e.g., $\mathcal{F}$=0.9), mR metrics show substantial improvement, highlighting **ImparTail**'s strong performance.
   (b) For SceneSayerSDE (Table 15, $\mathcal{F}$=0.9), mR@10 increases from 19.10 to 29.30 (+53.40%) with **ImparTail**.
   (c) **ImparTail** achieves balanced gains.For example in Table 13 ($\mathcal{F}$=0.5), for DSGDet++, **ImparTail** improves R@10 from 21.9 to 22.8 (+4.11%) and mR@10 from 11.80 to 13.50 (+14.41%).
4. Table 16, Table 17, Table 18, Table 19, present the With Constraint evaluation results for Scene Graph Generation (SGA) for GAGS. Table 20, Table 21, Table 22, Table 23, present the With Constraint evaluation results for Scene Graph Generation (SGA) for PGAGS. Table 24, Table 25, Table 26, Table 27 present the With Constraint evaluation results for Scene Graph Generation (SGA) for GAGS.
   (a) **ImparTail** consistently improves under constrained settings but with smaller gains than the No Constraint scenario.
   (b) GAGS (Tables 16-19) for fully grounded relationships and constraints result in consistent performance gains with **ImparTail**, especially for SceneSayerSDE.
   (c) Gains in mR metrics dominate, with the highest improvements observed at F=0.9. In PGAGS, gains are moderate compared to GAGS, with mR metrics seeing smaller improvements.

6.2.2. Results - No Constraint Setting - Grounded Action Genome Scenes (GAGS)

Table 4. **GAGS**-No Constraint-0.3 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.3	STTran++ [30]	-	39.30	55.60	**65.20**	**65.80**	23.10	35.50	59.20	**70.60**
	+ImparTail (Ours)	70	**44.70** (+13.74%)	**57.60** (+3.60%)	65.00 (-0.31%)	65.80	**24.60** (+6.49%)	**37.60** (+5.92%)	59.50 (+0.51%)	70.60
	+ImparTail (Ours)	40	40.20 (+2.29%)	55.10 (-0.90%)	64.80 (-0.61%)	65.80	21.50 (-6.93%)	33.50 (-5.63%)	56.20 (-5.07%)	69.80 (-1.13%)
	+ImparTail (Ours)	10	40.50 (+3.05%)	55.20 (-0.72%)	64.60 (-0.92%)	65.70 (-0.15%)	21.70 (-6.06%)	35.70 (+0.56%)	**59.90** (+1.18%)	70.60
	DSGDetr++ [30]	-	43.80	**57.80**	**65.10**	**65.80**	21.20	34.30	57.40	70.10
	+ImparTail (Ours)	70	44.10 (+0.68%)	57.30 (-0.87%)	65.10	65.80	22.10 (+4.25%)	35.60 (+3.79%)	59.50 (+3.66%)	69.90 (-0.29%)
	+ImparTail (Ours)	40	**44.30** (+1.14%)	57.40 (-0.69%)	65.10	65.80	**26.20** (+23.58%)	**39.90** (+16.33%)	**60.00** (+4.53%)	69.90 (-0.29%)
	+ImparTail (Ours)	10	37.50 (-14.38%)	52.90 (-8.48%)	63.90 (-1.84%)	65.70 (-0.15%)	20.10 (-5.19%)	33.50 (-2.33%)	58.40 (+1.74%)	**70.80** (+1.00%)
	SceneSayerODE [30]	-	40.30	54.00	63.80	**65.70**	22.20	34.50	56.70	68.20
	+ImparTail (Ours)	70	**40.70** (+0.99%)	**55.70** (+3.15%)	**64.70** (+1.41%)	65.70	**25.10** (+13.06%)	**38.40** (+11.30%)	**60.30** (+6.35%)	**69.70** (+2.20%)
	+ImparTail (Ours)	40	39.70 (-1.49%)	53.80 (-0.37%)	63.80	65.60 (-0.15%)	20.00 (-9.91%)	31.50 (-8.70%)	54.40 (-4.06%)	67.80 (-0.59%)
	+ImparTail (Ours)	10	28.70 (-28.78%)	44.60 (-17.41%)	61.50 (-3.61%)	65.40 (-0.46%)	22.80 (+2.70%)	35.90 (+4.06%)	57.40 (+1.23%)	69.50 (+1.91%)
	SceneSayerSDE [30]	-	**46.40**	58.80	**65.20**	**65.80**	23.10	35.70	57.70	68.60
	+ImparTail (Ours)	70	44.60 (-3.88%)	58.20 (-1.02%)	65.20	65.80	28.80 (+24.68%)	43.70 (+22.41%)	62.80 (+8.84%)	**71.20** (+3.79%)
	+ImparTail (Ours)	40	46.10 (-0.65%)	**58.90** (+0.17%)	65.20	65.80	24.90 (+7.79%)	37.80 (+5.88%)	58.60 (+1.56%)	68.90 (+0.44%)
	+ImparTail (Ours)	10	38.90 (-16.16%)	54.00 (-8.16%)	64.80 (-0.61%)	65.80	**31.60** (+36.80%)	**44.60** (+24.93%)	**64.50** (+11.79%)	71.20 (+3.79%)

Table 5. **GAGS**-No Constraint-0.5 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.5	STTran++ [30]	-	45.10	63.30	**73.40**	**74.00**	25.20	39.40	63.50	75.30
	+ImparTail (Ours)	70	**50.50** (+11.97%)	**64.90** (+2.53%)	73.30 (-0.14%)	74.00	**27.30** (+8.33%)	**41.60** (+5.58%)	**64.50** (+1.57%)	**75.40** (+0.13%)
	+ImparTail (Ours)	40	45.60 (+1.11%)	62.60 (-1.11%)	73.00 (-0.54%)	74.00	23.90 (-5.16%)	37.40 (-5.08%)	60.70 (-4.41%)	74.90 (-0.53%)
	+ImparTail (Ours)	10	45.40 (+0.67%)	61.90 (-2.21%)	72.60 (-1.09%)	73.90 (-0.14%)	23.70 (-5.95%)	39.20 (-0.51%)	64.10 (+0.94%)	75.00 (-0.40%)
	DSGDetr++ [30]	-	49.50	**65.40**	**73.40**	**74.00**	23.20	37.30	62.90	74.90
	+ImparTail (Ours)	70	**50.10** (+1.21%)	65.00 (-0.61%)	73.40	74.00	24.90 (+7.33%)	39.70 (+6.43%)	64.10 (+1.91%)	75.00 (+0.13%)
	+ImparTail (Ours)	40	49.90 (+0.81%)	64.50 (-1.38%)	73.40	74.00	**28.20** (+21.55%)	**42.20** (+13.14%)	**64.90** (+3.18%)	**75.20** (+0.40%)
	+ImparTail (Ours)	10	41.70 (-15.76%)	60.00 (-8.26%)	72.30 (-1.50%)	73.90 (-0.14%)	23.30 (+0.43%)	38.40 (+2.95%)	63.20 (+0.48%)	75.20 (+0.40%)
	SceneSayerODE [30]	-	**47.20**	62.40	72.50	73.90	24.90	38.00	61.80	74.30
	+ImparTail (Ours)	70	46.80 (-0.85%)	**63.30** (+1.44%)	**73.00** (+0.69%)	**74.00** (+0.14%)	**27.90** (+12.05%)	**42.50** (+11.84%)	**65.20** (+5.50%)	**75.40** (+1.48%)
	+ImparTail (Ours)	40	46.60 (-1.27%)	62.30 (-0.16%)	72.30 (-0.28%)	73.90	22.00 (-11.65%)	35.10 (-7.63%)	60.50 (-2.10%)	73.60 (-0.94%)
	+ImparTail (Ours)	10	34.50 (-26.91%)	52.70 (-15.54%)	70.40 (-2.90%)	73.80 (-0.14%)	26.50 (+6.43%)	40.90 (+7.63%)	63.20 (+2.27%)	74.90 (+0.81%)
	SceneSayerSDE [30]	-	**52.00**	**66.20**	73.40	**74.00**	25.00	39.00	62.70	73.70
	+ImparTail (Ours)	70	50.10 (-3.65%)	65.30 (-1.36%)	**73.50** (+0.14%)	74.00	31.80 (+27.20%)	46.70 (+19.74%)	67.50 (+7.66%)	**75.40** (+2.31%)
	+ImparTail (Ours)	40	51.80 (-0.38%)	66.20	73.40	74.00	27.00 (+8.00%)	40.70 (+4.36%)	63.60 (+1.44%)	74.80 (+1.49%)
	+ImparTail (Ours)	10	44.00 (-15.38%)	61.10 (-7.70%)	73.00 (-0.54%)	74.00	**34.60** (+38.40%)	**49.00** (+25.64%)	**68.50** (+9.25%)	75.30 (+2.17%)

Table 6. **GAGS**-No Constraint-0.7 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	**R@20**	**R@50**	**R@100**	**mR@10**	**mR@20**	**mR@50**	**mR@100**
0.7	STTran++ [30]	-	54.70	74.20	**83.40**	**83.80**	31.20	47.00	75.40	86.00
	+ImpartTail (Ours)	70	**60.40** (+10.42%)	**76.10** (+2.56%)	83.40	83.80	**34.40** (+10.26%)	**51.50** (+9.57%)	76.10 (+0.93%)	86.00
	+ImpartTail (Ours)	40	56.40 (+3.11%)	73.60 (-0.81%)	83.10 (-0.36%)	83.80	30.70 (-1.60%)	46.50 (-1.06%)	73.30 (-2.79%)	85.90 (-0.12%)
	+ImpartTail (Ours)	10	54.40 (-0.55%)	72.60 (-2.16%)	82.90 (-0.60%)	83.80	31.10 (-0.32%)	49.20 (+4.68%)	**76.50** (+1.46%)	**86.20** (+0.23%)
	DSGDetr++ [30]	-	59.60	**76.10**	**83.40**	**83.80**	28.60	46.10	73.80	85.80
	+ImpartTail (Ours)	70	**60.20** (+1.01%)	75.80 (-0.39%)	83.40	83.80	32.40 (+13.29%)	49.70 (+7.81%)	76.00 (+2.98%)	86.00 (+0.23%)
	+ImpartTail (Ours)	40	59.90 (+0.50%)	75.50 (-0.79%)	83.40	83.80	**35.70** (+24.83%)	**53.20** (+15.40%)	**76.50** (+3.66%)	85.80
	+ImpartTail (Ours)	10	50.40 (-15.44%)	70.90 (-6.83%)	82.80 (-0.72%)	83.80	31.40 (+9.79%)	49.70 (+7.81%)	75.70 (+2.57%)	**86.10** (+0.35%)
	SceneSayerODE [30]	-	**58.50**	74.00	82.80	**83.80**	29.80	45.20	72.00	84.20
	+ImpartTail (Ours)	70	56.80 (-2.91%)	**74.60** (+0.81%)	**83.10** (+0.36%)	83.80	**32.70** (+9.73%)	**51.10** (+13.05%)	**75.30** (+4.58%)	84.90 (+0.83%)
	+ImpartTail (Ours)	40	57.90 (-1.03%)	74.00	82.80	83.80	27.40 (-8.05%)	44.60 (-1.33%)	71.80 (-0.28%)	84.10 (-0.12%)
	+ImpartTail (Ours)	10	44.90 (-23.25%)	65.10 (-12.03%)	81.50 (-1.57%)	83.80	31.70 (+6.38%)	49.90 (+10.40%)	73.80 (+2.50%)	**85.90** (+2.02%)
	SceneSayerSDE [30]	-	**61.40**	**76.20**	83.30	**83.80**	30.20	45.40	72.80	84.00
	+ImpartTail (Ours)	70	60.20 (-1.95%)	76.10 (-0.13%)	**83.40** (+0.12%)	83.80	37.90 (+25.50%)	55.60 (+22.47%)	77.30 (+6.18%)	86.10 (+2.50%)
	+ImpartTail (Ours)	40	61.20 (-0.33%)	76.20	83.20 (-0.12%)	83.80	31.60 (+4.64%)	47.90 (+5.51%)	73.40 (+0.82%)	85.20 (+1.43%)
	+ImpartTail (Ours)	10	53.30 (-13.19%)	72.30 (-5.12%)	83.10 (-0.24%)	83.80	**40.90** (+35.43%)	**58.10** (+27.97%)	**78.30** (+7.55%)	**86.20** (+2.62%)

Table 7. **GAGS**-No Constraint-0.9 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	**R@20**	**R@50**	**R@100**	**mR@10**	**mR@20**	**mR@50**	**mR@100**
0.9	STTran++ [30]	-	68.70	86.80	**93.50**	**93.80**	42.50	60.80	84.80	**94.90**
	+ImpartTail (Ours)	70	**74.30** (+8.15%)	**88.00** (+1.38%)	93.50	93.80	**47.50** (+11.76%)	**64.70** (+6.41%)	89.00 (+4.95%)	94.90
	+ImpartTail (Ours)	40	70.50 (+2.62%)	86.20 (-0.69%)	93.50	93.80	43.00 (+1.18%)	61.00 (+0.33%)	88.50 (+4.36%)	94.70 (-0.21%)
	+ImpartTail (Ours)	10	67.50 (-1.75%)	85.20 (-1.84%)	93.40 (-0.11%)	93.80	46.30 (+8.94%)	64.40 (+5.92%)	**91.40** (+7.78%)	94.90
	DSGDetr++ [30]	-	73.60	**88.30**	**93.50**	**93.80**	39.00	58.80	83.20	**94.90**
	+ImpartTail (Ours)	70	**74.10** (+0.68%)	87.90 (-0.45%)	93.50	93.80	46.30 (+18.72%)	63.50 (+7.99%)	88.90 (+6.85%)	94.90
	+ImpartTail (Ours)	40	73.10 (-0.68%)	87.40 (-1.02%)	93.50	93.80	**50.00** (+28.21%)	**67.60** (+14.97%)	88.80 (+6.73%)	94.70 (-0.21%)
	+ImpartTail (Ours)	10	63.90 (-13.18%)	83.90 (-4.98%)	93.30 (-0.21%)	93.80	46.90 (+20.26%)	65.70 (+11.73%)	**91.00** (+9.38%)	94.80 (-0.11%)
	SceneSayerODE [30]	-	**73.00**	**87.20**	**93.30**	**93.80**	37.20	54.30	81.50	94.70
	+ImpartTail (Ours)	70	70.50 (-3.42%)	86.40 (-0.92%)	93.30	93.80	**40.80** (+9.68%)	57.80 (+6.45%)	84.70 (+3.93%)	94.80 (+0.11%)
	+ImpartTail (Ours)	40	72.50 (-0.68%)	86.70 (-0.57%)	93.20 (-0.11%)	93.80	34.40 (-7.53%)	52.10 (-4.05%)	83.60 (+2.58%)	94.40 (-0.32%)
	+ImpartTail (Ours)	10	61.60 (-15.62%)	81.40 (-6.65%)	93.00 (-0.32%)	93.80	39.90 (+7.26%)	**58.60** (+7.92%)	**87.00** (+6.75%)	**94.90** (+0.21%)
	SceneSayerSDE [30]	-	**73.70**	87.30	93.30	**93.80**	37.30	54.00	80.50	94.70
	+ImpartTail (Ours)	70	73.00 (-0.95%)	**87.40** (+0.11%)	**93.50** (+0.21%)	93.80	45.50 (+21.98%)	63.10 (+16.85%)	89.80 (+11.55%)	**94.90** (+0.21%)
	+ImpartTail (Ours)	40	73.30 (-0.54%)	87.30	93.40 (+0.11%)	93.80	38.70 (+3.75%)	56.50 (+4.63%)	82.30 (+2.24%)	94.80 (+0.11%)
	+ImpartTail (Ours)	10	67.00 (-9.09%)	84.80 (-2.86%)	93.30	93.80	**49.00** (+31.37%)	**67.30** (+24.63%)	**91.40** (+13.54%)	94.90 (+0.21%)

6.2.3. Results - No Constraint Setting - Partially Grounded Action Genome Scenes (PGAGS)

Table 8. **PGAGS**-No Constraint-0.3 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.3	STTran++ [30]	-	31.00	42.30	46.60	46.80	18.50	28.70	48.80	52.50
	+ImparTail (Ours)	70	33.20 (+7.10%)	41.30 (-2.36%)	46.60	**47.00** (+0.43%)	16.60 (-10.27%)	27.20 (-5.23%)	49.30 (+1.02%)	**52.80** (+0.57%)
	+ImparTail (Ours)	40	**34.00** (+9.68%)	**42.50** (+0.47%)	**46.70** (+0.21%)	47.00 (+0.43%)	19.70 (+6.49%)	29.60 (+3.14%)	**49.60** (+1.64%)	52.80 (+0.57%)
	+ImparTail (Ours)	10	29.40 (-5.16%)	40.20 (-4.96%)	45.80 (-1.72%)	46.20 (-1.28%)	**21.20** (+14.59%)	**31.50** (+9.76%)	49.20 (+0.82%)	52.40 (-0.19%)
	DSGDetr++ [30]	-	**35.60**	**44.00**	**48.20**	**48.50**	17.90	28.60	49.20	**53.80**
	+ImparTail (Ours)	70	34.40 (-3.37%)	43.50 (-1.14%)	48.20	48.50	**21.10** (+17.88%)	30.50 (+6.64%)	49.70 (+1.02%)	53.80
	+ImparTail (Ours)	40	32.20 (-9.55%)	41.80 (-5.00%)	47.50 (-1.45%)	47.80 (-1.44%)	18.10 (+1.12%)	28.40 (-0.70%)	49.00 (-0.41%)	52.80 (-1.86%)
	+ImparTail (Ours)	10	32.20 (-9.55%)	42.00 (-4.55%)	47.70 (-1.04%)	48.00 (-1.03%)	21.00 (+17.32%)	**31.60** (+10.49%)	**50.80** (+3.25%)	53.50 (-0.56%)
	SceneSayerODE [30]	-	30.00	39.80	46.90	48.10	14.30	23.30	44.60	53.50
	+ImparTail (Ours)	70	25.20 (-16.00%)	35.70 (-10.30%)	45.20 (-3.62%)	47.30 (-1.66%)	16.90 (+18.18%)	25.70 (+10.30%)	45.60 (+2.24%)	52.60 (-1.68%)
	+ImparTail (Ours)	40	**34.80** (+16.00%)	**42.40** (+6.53%)	**47.00** (+0.21%)	47.70 (-0.83%)	20.40 (+42.66%)	29.90 (+28.33%)	47.30 (+6.05%)	53.40 (-0.19%)
	+ImparTail (Ours)	10	27.70 (-7.67%)	37.20 (-6.53%)	46.10 (-1.71%)	**48.20** (+0.21%)	**21.90** (+53.15%)	**31.40** (+34.76%)	**49.30** (+10.54%)	**53.70** (+0.37%)
	SceneSayerSDE [30]	-	**35.90**	**43.70**	**47.80**	**48.40**	20.60	31.20	49.80	**55.00**
	+ImparTail (Ours)	70	34.80 (-3.06%)	42.80 (-2.06%)	47.10 (-1.46%)	48.00 (-0.83%)	24.30 (+17.96%)	34.50 (+10.58%)	50.80 (+2.01%)	54.20 (-1.45%)
	+ImparTail (Ours)	40	35.50 (-1.11%)	43.20 (-1.14%)	47.60 (-0.42%)	48.20 (-0.41%)	21.50 (+4.37%)	32.10 (+2.88%)	49.10 (-1.41%)	54.40 (-1.09%)
	+ImparTail (Ours)	10	33.30 (-7.24%)	41.90 (-4.12%)	47.00 (-1.67%)	47.90 (-1.03%)	**27.50** (+33.50%)	**37.70** (+20.83%)	**52.60** (+5.62%)	54.50 (-0.91%)

Table 9. **PGAGS**-No Constraint-0.5 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.5	STTran++ [30]	-	34.30	46.70	**51.50**	**51.60**	20.90	32.50	50.10	**53.20**
	+ImparTail (Ours)	70	36.60 (+6.71%)	45.70 (-2.14%)	51.10 (-0.78%)	51.50 (-0.19%)	18.70 (-10.53%)	30.70 (-5.54%)	49.40 (-1.40%)	52.70 (-0.94%)
	+ImparTail (Ours)	40	**37.30** (+8.75%)	**46.90** (+0.43%)	51.10 (-0.78%)	51.30 (-0.58%)	21.90 (+4.78%)	33.20 (+2.15%)	50.00 (-0.20%)	52.30 (-1.69%)
	+ImparTail (Ours)	10	32.50 (-5.25%)	44.10 (-5.57%)	50.40 (-2.14%)	50.80 (-1.55%)	**24.00** (+14.83%)	**34.90** (+7.38%)	**50.50** (+0.80%)	52.70 (-0.94%)
	DSGDetr++ [30]	-	**39.60**	**49.60**	**54.30**	**54.50**	19.90	32.30	50.60	55.80
	+ImparTail (Ours)	70	37.40 (-5.56%)	48.30 (-2.62%)	53.70 (-1.10%)	54.00 (-0.92%)	**23.10** (+16.08%)	33.60 (+4.02%)	50.50 (-0.20%)	55.30 (-0.90%)
	+ImparTail (Ours)	40	36.30 (-8.33%)	47.10 (-5.04%)	53.00 (-2.39%)	53.30 (-2.20%)	20.50 (+3.02%)	31.60 (-2.17%)	50.00 (-1.19%)	55.00 (-1.43%)
	+ImparTail (Ours)	10	35.50 (-10.35%)	47.30 (-4.64%)	54.10 (-0.37%)	54.50	22.60 (+13.57%)	**35.20** (+8.98%)	**52.50** (+3.75%)	**56.70** (+1.61%)
	SceneSayerODE [30]	-	34.10	45.10	52.80	54.10	16.90	26.30	45.70	54.00
	+ImparTail (Ours)	70	29.00 (-14.96%)	41.40 (-8.20%)	51.90 (-1.70%)	54.10	19.70 (+16.57%)	29.70 (+12.93%)	47.40 (+3.72%)	54.40 (+0.74%)
	+ImparTail (Ours)	40	**39.10** (+14.66%)	**48.00** (+6.43%)	**53.40** (+1.14%)	**54.20** (+0.18%)	22.70 (+34.32%)	33.60 (+27.76%)	49.70 (+8.75%)	**55.60** (+2.96%)
	+ImparTail (Ours)	10	32.10 (-5.87%)	42.90 (-4.88%)	52.20 (-1.14%)	54.10	**25.00** (+47.93%)	**36.00** (+36.88%)	**50.40** (+10.28%)	55.40 (+2.59%)
	SceneSayerSDE [30]	-	**39.90**	**48.90**	**53.70**	**54.50**	22.90	34.30	51.00	56.20
	+ImparTail (Ours)	70	38.70 (-3.01%)	47.90 (-2.04%)	53.30 (-0.74%)	54.00 (-0.92%)	26.90 (+17.47%)	37.70 (+9.91%)	**52.70** (+3.33%)	**56.70** (+0.89%)
	+ImparTail (Ours)	40	39.70 (-0.50%)	48.40 (-1.02%)	53.50 (-0.37%)	54.20 (-0.55%)	23.60 (+3.06%)	34.30	50.60 (-0.78%)	56.00 (-0.36%)
	+ImparTail (Ours)	10	37.00 (-7.27%)	46.80 (-4.29%)	52.80 (-1.68%)	53.90 (-1.10%)	**30.30** (+32.31%)	**40.90** (+19.24%)	52.30 (+2.55%)	55.00 (-2.14%)

CVPR
#16783

Table 10. **PGAGS**-No Constraint-0.7 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.7	STTran++ [30]	-	41.40	54.90	**59.60**	**59.70**	25.30	38.30	56.10	58.40
	+ImparTail (Ours)	70	44.80 (+8.21%)	54.40 (-0.91%)	59.30 (-0.50%)	59.50 (-0.34%)	24.40 (-3.56%)	38.70 (+1.04%)	56.40 (+0.53%)	**59.90** (+2.57%)
	+ImparTail (Ours)	40	**45.50** (+9.90%)	**55.50** (+1.09%)	59.60	59.70	28.00 (+10.67%)	41.70 (+8.88%)	57.10 (+1.78%)	59.30 (+1.54%)
	+ImparTail (Ours)	10	39.60 (-4.35%)	53.00 (-3.46%)	59.10 (-0.84%)	59.50 (-0.34%)	**29.80** (+17.79%)	**43.90** (+14.62%)	**57.60** (+2.67%)	59.40 (+1.71%)
	DSGDetr++ [30]	-	**47.70**	**58.10**	**62.40**	**62.60**	24.80	39.50	57.30	61.10
	+ImparTail (Ours)	70	45.40 (-4.82%)	57.30 (-1.38%)	62.10 (-0.48%)	62.20 (-0.64%)	**28.70** (+15.73%)	40.90 (+3.54%)	57.60 (+0.52%)	60.80 (-0.49%)
	+ImparTail (Ours)	40	44.30 (-7.13%)	55.90 (-3.79%)	61.20 (-1.92%)	61.40 (-1.92%)	26.30 (+6.05%)	41.00 (+3.80%)	58.10 (+1.40%)	**61.90** (+1.31%)
	+ImparTail (Ours)	10	42.80 (-10.27%)	55.60 (-4.30%)	62.00 (-0.64%)	62.30 (-0.48%)	28.10 (+13.31%)	**42.20** (+6.84%)	**58.50** (+2.09%)	61.80 (+1.15%)
	SceneSayerODE [30]	-	40.90	53.00	60.40	61.60	20.50	32.40	52.80	60.10
	+ImparTail (Ours)	70	35.30 (-13.69%)	49.00 (-7.55%)	59.40 (-1.66%)	61.30 (-0.49%)	23.40 (+14.15%)	36.10 (+11.42%)	52.70 (-0.19%)	59.60 (-0.83%)
	+ImparTail (Ours)	40	**46.00** (+12.47%)	**56.00** (+5.66%)	**61.20** (+1.32%)	61.90 (+0.49%)	27.10 (+32.20%)	40.00 (+23.46%)	**56.40** (+6.82%)	**61.20** (+1.83%)
	+ImparTail (Ours)	10	40.40 (-1.22%)	52.20 (-1.51%)	60.60 (+0.33%)	**62.00** (+0.65%)	**30.70** (+49.76%)	**43.90** (+35.49%)	55.90 (+5.87%)	60.20 (+0.17%)
	SceneSayerSDE [30]	-	**47.40**	**57.00**	61.90	62.50	27.00	40.20	57.20	61.70
	+ImparTail (Ours)	70	46.50 (-1.90%)	56.10 (-1.58%)	61.30 (-0.97%)	62.10 (-0.64%)	32.10 (+18.89%)	45.20 (+12.44%)	**60.40** (+5.59%)	**63.50** (+2.92%)
	+ImparTail (Ours)	40	47.30 (-0.21%)	56.90 (-0.18%)	**62.00** (+0.16%)	**62.60** (+0.16%)	28.40 (+5.19%)	40.50 (+0.75%)	57.90 (+1.22%)	63.50 (+2.92%)
	+ImparTail (Ours)	10	44.30 (-6.54%)	55.20 (-3.16%)	61.00 (-1.45%)	62.00 (-0.80%)	**35.50** (+31.48%)	**48.20** (+19.90%)	58.50 (+2.27%)	61.20 (-0.81%)

Table 11. **PGAGS**-No Constraint-0.9 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.9	STTran++ [30]	-	50.00	61.60	64.40	64.50	32.40	46.90	60.60	62.60
	+ImparTail (Ours)	70	**54.00** (+8.00%)	**62.30** (+1.14%)	**65.00** (+0.93%)	65.00 (+0.78%)	35.10 (+8.33%)	48.70 (+3.84%)	61.40 (+1.32%)	63.10 (+0.80%)
	+ImparTail (Ours)	40	53.30 (+6.60%)	62.20 (+0.97%)	65.00 (+0.93%)	**65.10** (+0.93%)	36.80 (+13.58%)	49.60 (+5.76%)	62.10 (+2.48%)	63.40 (+1.28%)
	+ImparTail (Ours)	10	48.20 (-3.60%)	60.50 (-1.79%)	64.90 (+0.78%)	65.10 (+0.93%)	**40.90** (+26.23%)	**52.10** (+11.09%)	**63.80** (+5.28%)	**64.60** (+3.19%)
	DSGDetr++ [30]	-	**57.10**	**66.40**	**69.40**	**69.50**	33.00	49.00	64.20	68.20
	+ImparTail (Ours)	70	54.90 (-3.85%)	65.40 (-1.51%)	68.80 (-0.86%)	68.90 (-0.86%)	35.90 (+8.79%)	49.40 (+0.82%)	63.90 (-0.47%)	67.50 (-1.03%)
	+ImparTail (Ours)	40	54.30 (-4.90%)	65.00 (-2.11%)	68.60 (-1.15%)	68.80 (-1.01%)	35.70 (+8.18%)	51.50 (+5.10%)	65.70 (+2.34%)	**69.30** (+1.61%)
	+ImparTail (Ours)	10	52.60 (-7.88%)	64.30 (-3.16%)	69.00 (-0.58%)	69.20 (-0.43%)	**38.30** (+16.06%)	**52.60** (+7.35%)	**67.30** (+4.83%)	69.20 (+1.47%)
	SceneSayerODE [30]	-	50.60	62.00	68.50	69.70	27.40	41.70	61.10	69.30
	+ImparTail (Ours)	70	44.80 (-11.46%)	58.50 (-5.65%)	68.30 (-0.29%)	70.00 (+0.43%)	29.90 (+9.12%)	42.60 (+2.16%)	60.80 (-0.49%)	68.50 (-1.15%)
	+ImparTail (Ours)	40	**54.70** (+8.10%)	**64.30** (+3.71%)	68.80 (+0.44%)	69.50 (-0.29%)	32.90 (+20.07%)	46.60 (+11.75%)	62.00 (+1.47%)	69.10 (-0.29%)
	+ImparTail (Ours)	10	51.20 (+1.19%)	63.00 (+1.61%)	**69.20** (+1.02%)	**70.30** (+0.86%)	**37.80** (+37.96%)	**52.90** (+26.86%)	**65.20** (+6.71%)	**69.60** (+0.43%)
	SceneSayerSDE [30]	-	**56.30**	**65.70**	69.50	70.00	33.20	48.20	64.30	69.50
	+ImparTail (Ours)	70	55.20 (-1.95%)	65.00 (-1.07%)	69.20 (-0.43%)	69.90 (-0.14%)	38.10 (+14.76%)	51.50 (+6.85%)	65.70 (+2.18%)	69.10 (-0.58%)
	+ImparTail (Ours)	40	56.30	65.70	**69.90** (+0.58%)	**70.60** (+0.86%)	34.10 (+2.71%)	48.10 (-0.21%)	65.70 (+2.18%)	**71.80** (+3.31%)
	+ImparTail (Ours)	10	53.70 (-4.62%)	64.20 (-2.28%)	69.10 (-0.58%)	70.20 (+0.29%)	**42.00** (+26.51%)	**54.50** (+13.07%)	**66.40** (+3.27%)	69.30 (-0.29%)

6.2.4. Results - No Constraint Setting - Action Genome Scenes (AGS)

Table 12. **AGS**-No Constraint-0.3 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.3	STTran++ [30]	-	**22.90**	**36.00**	**51.30**	**55.20**	**13.10**	20.20	36.20	49.90
	+ImparTail (Ours)	70	22.70 (-0.87%)	35.80 (-0.56%)	50.50 (-1.56%)	55.20	12.20 (-6.87%)	19.00 (-5.94%)	35.70 (-1.38%)	50.30 (+0.80%)
	+ImparTail (Ours)	40	22.10 (-3.49%)	35.50 (-1.39%)	50.50 (-1.56%)	55.00 (-0.36%)	12.80 (-2.29%)	**21.10** (+4.46%)	**37.50** (+3.59%)	**51.10** (+2.40%)
	+ImparTail (Ours)	10	19.50 (-14.85%)	30.40 (-15.56%)	48.50 (-5.46%)	54.90 (-0.54%)	11.10 (-15.27%)	18.50 (-8.42%)	36.50 (+0.83%)	51.10 (+2.40%)
	DSGDetr++ [30]	-	20.80	33.10	**49.20**	53.90	11.50	17.80	34.10	49.10
	+ImparTail (Ours)	70	**21.30** (+2.40%)	**33.60** (+1.51%)	49.00 (-0.41%)	**54.30** (+0.74%)	12.40 (+7.83%)	19.30 (+8.43%)	34.90 (+2.35%)	48.90 (-0.41%)
	+ImparTail (Ours)	40	18.40 (-11.54%)	29.70 (-10.27%)	47.30 (-3.86%)	53.80 (-0.19%)	12.50 (+8.70%)	20.60 (+15.73%)	36.50 (+7.04%)	49.20 (+0.20%)
	+ImparTail (Ours)	10	18.90 (-9.13%)	29.90 (-9.67%)	47.10 (-4.27%)	53.10 (-1.48%)	**12.80** (+11.30%)	**21.80** (+22.47%)	**37.50** (+9.97%)	**49.70** (+1.22%)
	SceneSayerODE [30]	-	22.60	31.60	44.50	51.70	12.60	19.30	32.70	44.30
	+ImparTail (Ours)	70	20.50 (-9.29%)	29.50 (-6.65%)	42.50 (-4.49%)	49.90 (-3.48%)	13.40 (+6.35%)	20.40 (+5.70%)	33.90 (+3.67%)	45.00 (+1.58%)
	+ImparTail (Ours)	40	**26.70** (+18.14%)	**36.10** (+14.24%)	**47.40** (+6.52%)	**52.60** (+1.74%)	**14.30** (+13.49%)	21.50 (+11.40%)	35.50 (+8.56%)	45.80 (+3.39%)
	+ImparTail (Ours)	10	22.60	32.10 (+1.58%)	45.20 (+1.57%)	52.20 (+0.97%)	14.20 (+12.70%)	**22.60** (+17.10%)	**37.50** (+14.68%)	**47.50** (+7.22%)
	SceneSayerSDE [30]	-	26.30	**35.70**	**47.20**	52.40	14.30	22.00	36.40	46.80
	+ImparTail (Ours)	70	25.90 (-1.52%)	35.30 (-1.12%)	47.20	**52.90** (+0.95%)	17.50 (+22.38%)	26.60 (+20.91%)	40.80 (+12.09%)	50.70 (+8.33%)
	+ImparTail (Ours)	40	**26.50** (+0.76%)	35.50 (-0.56%)	47.20	52.40	15.30 (+6.99%)	22.90 (+4.09%)	37.10 (+1.92%)	47.60 (+1.71%)
	+ImparTail (Ours)	10	22.80 (-13.31%)	32.20 (-9.80%)	44.60 (-5.51%)	51.20 (-2.29%)	**20.20** (+41.26%)	**29.30** (+33.18%)	**43.30** (+18.96%)	**51.30** (+9.62%)

Table 13. **AGS**-No Constraint-0.5 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.5	STTran++ [30]	-	**24.60**	**38.40**	**54.50**	**58.60**	**13.90**	**21.30**	38.50	52.50
	+ImparTail (Ours)	70	24.00 (-2.44%)	38.10 (-0.78%)	54.00 (-0.92%)	58.60	12.60 (-9.35%)	19.90 (-6.57%)	37.90 (-1.56%)	53.40 (+1.71%)
	+ImparTail (Ours)	40	23.10 (-6.10%)	37.60 (-2.08%)	53.60 (-1.65%)	58.50 (-0.17%)	12.90 (-7.19%)	21.30	39.00 (+1.30%)	53.90 (+2.67%)
	+ImparTail (Ours)	10	21.00 (-14.63%)	33.10 (-13.80%)	51.90 (-4.77%)	58.40 (-0.34%)	12.10 (-12.95%)	20.00 (-6.10%)	**39.60** (+2.86%)	**54.30** (+3.43%)
	DSGDetr++ [30]	-	21.90	34.70	52.10	57.30	11.80	18.20	36.10	52.70
	+ImparTail (Ours)	70	**22.80** (+4.11%)	**36.10** (+4.03%)	**52.90** (+1.54%)	**58.30** (+1.75%)	13.10 (+11.02%)	20.60 (+13.19%)	37.80 (+4.71%)	55.60 (+5.50%)
	+ImparTail (Ours)	40	20.00 (-8.68%)	32.30 (-6.92%)	51.30 (-1.54%)	57.50 (+0.35%)	13.10 (+11.02%)	21.60 (+18.68%)	39.60 (+9.70%)	**56.40** (+7.02%)
	+ImparTail (Ours)	10	20.00 (-8.68%)	32.10 (-7.49%)	50.50 (-3.07%)	56.70 (-1.05%)	**13.50** (+14.41%)	**22.90** (+25.82%)	**40.70** (+12.74%)	52.60 (-0.19%)
	SceneSayerODE [30]	-	25.80	35.80	48.90	55.00	14.00	22.30	36.50	49.00
	+ImparTail (Ours)	70	23.40 (-9.30%)	33.50 (-6.42%)	47.00 (-3.89%)	54.20 (-1.45%)	14.90 (+6.43%)	23.20 (+4.04%)	37.30 (+2.19%)	48.40 (-1.22%)
	+ImparTail (Ours)	40	**29.40** (+13.95%)	**39.50** (+10.34%)	**51.40** (+5.11%)	**56.10** (+2.00%)	15.30 (+9.29%)	22.90 (+2.69%)	37.90 (+3.84%)	49.00
	+ImparTail (Ours)	10	24.90 (-3.49%)	34.90 (-2.51%)	48.60 (-0.61%)	55.60 (+1.09%)	**15.60** (+11.43%)	**24.80** (+11.21%)	**39.70** (+8.77%)	**50.50** (+3.06%)
	SceneSayerSDE [30]	-	29.00	39.10	51.50	**56.70**	15.40	23.70	38.70	50.80
	+ImparTail (Ours)	70	28.10 (-3.10%)	38.20 (-2.30%)	50.90 (-1.17%)	56.40 (-0.53%)	18.20 (+18.18%)	27.10 (+14.35%)	42.80 (+10.59%)	52.80 (+3.94%)
	+ImparTail (Ours)	40	**29.20** (+0.69%)	**39.30** (+0.51%)	**51.80** (+0.58%)	56.60 (-0.18%)	16.50 (+7.14%)	25.10 (+5.91%)	40.10 (+3.62%)	50.60 (-0.39%)
	+ImparTail (Ours)	10	25.10 (-13.45%)	35.70 (-8.70%)	48.50 (-5.83%)	54.90 (-3.17%)	**21.80** (+41.56%)	**31.40** (+32.49%)	**45.40** (+17.31%)	**53.80** (+5.91%)

CVPR
#16783

Table 14. **AGS**-No Constraint-0.7 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.7	STTran++ [30]	-	**27.20**	42.10	**59.70**	**63.90**	**15.70**	23.70	41.90	57.50
	+ImparTail (Ours)	70	26.80 (-1.47%)	**43.00** (+2.14%)	59.70	63.90	14.90 (-5.10%)	23.90 (+0.84%)	42.60 (+1.67%)	58.30 (+1.39%)
	+ImparTail (Ours)	40	25.80 (-5.15%)	41.80 (-0.71%)	59.30 (-0.67%)	63.80 (-0.16%)	14.90 (-5.10%)	**24.70** (+4.22%)	**44.00** (+5.01%)	58.90 (+2.43%)
	+ImparTail (Ours)	10	23.60 (-13.24%)	36.50 (-13.30%)	57.00 (-4.52%)	63.50 (-0.63%)	14.00 (-10.83%)	23.20 (-2.11%)	43.70 (+4.30%)	**59.60** (+3.65%)
	DSGDetr++ [30]	-	**24.90**	38.20	56.90	62.00	13.20	20.00	38.80	56.50
	+ImparTail (Ours)	70	24.80 (-0.40%)	**40.00** (+4.71%)	**57.80** (+1.58%)	**63.20** (+1.94%)	14.40 (+9.09%)	23.20 (+16.00%)	41.30 (+6.44%)	56.90 (+0.71%)
	+ImparTail (Ours)	40	22.50 (-9.64%)	36.70 (-3.93%)	56.30 (-1.05%)	61.70 (-0.48%)	14.90 (+12.88%)	24.80 (+24.00%)	43.90 (+13.14%)	57.20 (+1.24%)
	+ImparTail (Ours)	10	22.20 (-10.84%)	35.70 (-6.54%)	56.00 (-1.58%)	62.10 (+0.16%)	**15.90** (+20.45%)	**26.70** (+33.50%)	**45.30** (+16.75%)	**57.90** (+2.48%)
	SceneSayerODE [30]	-	31.70	42.10	55.10	60.50	16.40	24.90	40.50	53.00
	+ImparTail (Ours)	70	28.90 (-8.83%)	39.50 (-6.18%)	53.30 (-3.27%)	59.90 (-0.99%)	17.80 (+8.54%)	27.00 (+8.43%)	41.60 (+2.72%)	53.40 (+0.75%)
	+ImparTail (Ours)	40	**34.00** (+7.26%)	**44.30** (+5.23%)	**56.00** (+1.63%)	**60.60** (+0.17%)	17.50 (+6.71%)	25.80 (+3.61%)	41.10 (+1.48%)	52.60 (-0.75%)
	+ImparTail (Ours)	10	29.20 (-7.89%)	40.00 (-4.99%)	53.40 (-3.09%)	60.00 (-0.83%)	**18.90** (+15.24%)	**29.40** (+18.07%)	**43.80** (+8.15%)	**53.90** (+1.70%)
	SceneSayerSDE [30]	-	33.80	43.90	56.40	61.10	17.30	26.10	42.50	54.00
	+ImparTail (Ours)	70	33.40 (-1.18%)	43.70 (-0.46%)	56.00 (-0.71%)	61.10	21.10 (+21.97%)	31.40 (+20.31%)	46.80 (+10.12%)	57.20 (+5.93%)
	+ImparTail (Ours)	40	**34.20** (+1.18%)	**44.40** (+1.14%)	**56.80** (+0.71%)	**61.40** (+0.49%)	19.00 (+9.83%)	27.80 (+6.51%)	43.60 (+2.59%)	54.40 (+0.74%)
	+ImparTail (Ours)	10	30.00 (-11.24%)	41.20 (-6.15%)	54.20 (-3.90%)	60.30 (-1.31%)	**25.60** (+47.98%)	**35.10** (+34.48%)	**50.00** (+17.65%)	**58.00** (+7.41%)

Table 15. **AGS**-No Constraint-0.9 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	NO CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.9	STTran++ [30]	-	**30.20**	45.80	64.40	**68.40**	18.20	27.50	47.20	62.80
	+ImparTail (Ours)	70	30.10 (-0.33%)	**47.80** (+4.37%)	**64.50** (+0.16%)	68.40	19.50 (+7.14%)	29.60 (+7.64%)	50.00 (+5.93%)	64.00 (+1.91%)
	+ImparTail (Ours)	40	29.30 (-2.98%)	47.10 (+2.84%)	64.50 (+0.16%)	68.30 (-0.15%)	**20.50** (+12.64%)	**31.80** (+15.64%)	**51.90** (+9.96%)	63.80 (+1.59%)
	+ImparTail (Ours)	10	26.30 (-12.91%)	40.70 (-11.14%)	60.90 (-5.43%)	68.00 (-0.58%)	18.80 (+3.30%)	30.40 (+10.55%)	51.40 (+8.90%)	**64.80** (+3.18%)
	DSGDetr++ [30]	-	**28.70**	43.40	62.70	67.10	15.50	23.80	44.90	61.70
	+ImparTail (Ours)	70	27.90 (-2.79%)	**45.10** (+3.92%)	**63.90** (+1.91%)	**68.40** (+1.94%)	18.50 (+19.35%)	28.70 (+20.59%)	48.50 (+8.02%)	62.00 (+0.49%)
	+ImparTail (Ours)	40	27.40 (-4.53%)	43.80 (+0.92%)	62.90 (+0.32%)	67.40 (+0.45%)	20.50 (+32.26%)	**32.30** (+35.71%)	**52.40** (+16.70%)	**63.00** (+2.11%)
	+ImparTail (Ours)	10	26.20 (-8.71%)	40.60 (-6.45%)	61.40 (-2.07%)	67.60 (+0.75%)	**21.20** (+36.77%)	32.10 (+34.87%)	51.20 (+14.03%)	61.70
	SceneSayerODE [30]	-	37.80	49.10	61.40	**66.20**	18.90	28.60	45.20	**58.40**
	+ImparTail (Ours)	70	35.60 (-5.82%)	47.00 (-4.28%)	60.00 (-2.28%)	65.20 (-1.51%)	21.20 (+12.17%)	31.80 (+11.19%)	48.20 (+6.64%)	58.20 (-0.34%)
	+ImparTail (Ours)	40	**39.30** (+3.97%)	**50.10** (+2.04%)	**61.60** (+0.33%)	66.10 (-0.15%)	20.10 (+6.35%)	29.60 (+3.50%)	45.60 (+0.88%)	57.80 (-1.03%)
	+ImparTail (Ours)	10	34.20 (-9.52%)	46.00 (-6.31%)	59.20 (-3.58%)	65.30 (-1.36%)	**23.10** (+22.22%)	**33.90** (+18.53%)	**49.50** (+9.51%)	58.00 (-0.68%)
	SceneSayerSDE [30]	-	38.50	49.90	61.90	66.50	19.10	29.00	45.60	59.40
	+ImparTail (Ours)	70	39.30 (+2.08%)	50.60 (+1.40%)	61.90	66.40 (-0.15%)	24.80 (+29.84%)	36.30 (+25.17%)	51.30 (+12.50%)	60.00 (+1.01%)
	+ImparTail (Ours)	40	**40.20** (+4.42%)	**50.80** (+1.80%)	**62.00** (+0.16%)	**66.70** (+0.30%)	21.80 (+14.14%)	31.40 (+8.28%)	47.60 (+4.39%)	58.70 (-1.18%)
	+ImparTail (Ours)	10	35.50 (-7.79%)	47.20 (-5.41%)	60.70 (-1.94%)	66.00 (-0.75%)	**29.30** (+53.40%)	**39.40** (+35.86%)	**53.00** (+16.23%)	**63.60** (+7.07%)

6.2.5. Results - With Constraint Setting - Grounded Action Genome Scenes (GAGS)

Table 16. **GAGS**-With Constraint-0.3 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	**R@20**	**R@50**	**R@100**	**mR@10**	**mR@20**	**mR@50**	**mR@100**
0.3	STTran++ [30]	-	26.50	29.50	29.60	29.60	15.90	18.50	18.60	18.60
	+ImpartTail (Ours)	70	**36.60** (+38.11%)	**39.70** (+34.58%)	**39.80** (+34.46%)	**39.80** (+34.46%)	**18.70** (+17.61%)	**21.40** (+15.68%)	**21.40** (+15.05%)	**21.40** (+15.05%)
	+ImpartTail (Ours)	40	32.60 (+23.02%)	35.60 (+20.68%)	35.60 (+20.27%)	35.60 (+20.27%)	17.00 (+6.92%)	19.40 (+4.86%)	19.40 (+4.30%)	19.40 (+4.30%)
	+ImpartTail (Ours)	10	33.30 (+25.66%)	35.80 (+21.36%)	35.80 (+20.95%)	35.80 (+20.95%)	17.80 (+11.95%)	20.40 (+10.27%)	20.40 (+9.68%)	20.40 (+9.68%)
	DSGDetr++ [30]	-	36.20	39.30	39.30	39.30	15.10	17.50	17.50	17.50
	+ImpartTail (Ours)	70	**37.00** (+2.21%)	**40.20** (+2.29%)	**40.30** (+2.54%)	**40.30** (+2.54%)	17.20 (+13.91%)	19.90 (+13.71%)	20.00 (+14.29%)	20.00 (+14.29%)
	+ImpartTail (Ours)	40	36.40 (+0.55%)	39.40 (+0.25%)	39.40 (+0.25%)	39.40 (+0.25%)	**18.90** (+25.17%)	**22.00** (+25.71%)	**22.00** (+25.71%)	**22.00** (+25.71%)
	+ImpartTail (Ours)	10	30.90 (-14.64%)	33.20 (-15.52%)	33.20 (-15.52%)	33.20 (-15.52%)	17.10 (+13.25%)	19.70 (+12.57%)	19.70 (+12.57%)	19.70 (+12.57%)
	SceneSayerODE [30]	-	**34.60**	**37.30**	**37.30**	**37.30**	15.20	17.90	18.00	18.00
	+ImpartTail (Ours)	70	33.20 (-4.05%)	36.00 (-3.49%)	36.00 (-3.49%)	36.00 (-3.49%)	**17.40** (+14.47%)	**21.10** (+17.88%)	**21.20** (+17.78%)	**21.20** (+17.78%)
	+ImpartTail (Ours)	40	33.70 (-2.60%)	36.40 (-2.41%)	36.40 (-2.41%)	36.40 (-2.41%)	13.30 (-12.50%)	15.90 (-11.17%)	15.90 (-11.67%)	15.90 (-11.67%)
	+ImpartTail (Ours)	10	23.10 (-33.24%)	24.70 (-33.78%)	24.70 (-33.78%)	24.70 (-33.78%)	16.50 (+8.55%)	19.80 (+10.61%)	19.90 (+10.56%)	19.90 (+10.56%)
	SceneSayerSDE [30]	-	38.30	41.70	41.70	41.70	16.10	19.20	19.30	19.30
	+ImpartTail (Ours)	70	36.40 (-4.96%)	39.70 (-4.80%)	39.80 (-4.56%)	39.80 (-4.56%)	20.30 (+26.09%)	24.60 (+28.12%)	24.70 (+27.98%)	24.70 (+27.98%)
	+ImpartTail (Ours)	40	**38.60** (+0.78%)	**42.00** (+0.72%)	**42.00** (+0.72%)	**42.00** (+0.72%)	16.10	19.20	19.20 (-0.52%)	19.20 (-0.52%)
	+ImpartTail (Ours)	10	30.60 (-20.10%)	33.00 (-20.86%)	33.00 (-20.86%)	33.00 (-20.86%)	**24.50** (+52.17%)	**29.90** (+55.73%)	**30.00** (+55.44%)	**30.00** (+55.44%)

Table 17. **GAGS**-With Constraint-0.5 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	**R@20**	**R@50**	**R@100**	**mR@10**	**mR@20**	**mR@50**	**mR@100**
0.5	STTran++ [30]	-	30.90	34.20	34.20	34.20	17.80	20.90	21.00	21.00
	+ImpartTail (Ours)	70	**41.70** (+34.95%)	**45.00** (+31.58%)	**45.10** (+31.87%)	**45.10** (+31.87%)	**21.30** (+19.66%)	**24.80** (+18.66%)	**24.80** (+18.10%)	**24.80** (+18.10%)
	+ImpartTail (Ours)	40	36.70 (+18.77%)	40.00 (+16.96%)	40.00 (+16.96%)	40.00 (+16.96%)	19.10 (+7.30%)	22.10 (+5.74%)	22.10 (+5.24%)	22.10 (+5.24%)
	+ImpartTail (Ours)	10	37.50 (+21.36%)	40.20 (+17.54%)	40.20 (+17.54%)	40.20 (+17.54%)	19.90 (+11.80%)	22.70 (+8.61%)	22.80 (+8.57%)	22.80 (+8.57%)
	DSGDetr++ [30]	-	41.20	44.60	44.70	44.70	17.10	20.00	20.00	20.00
	+ImpartTail (Ours)	70	**41.90** (+1.70%)	**45.40** (+1.79%)	**45.40** (+1.57%)	**45.40** (+1.57%)	19.90 (+16.37%)	23.10 (+15.50%)	23.10 (+15.50%)	23.10 (+15.50%)
	+ImpartTail (Ours)	40	41.20	44.50 (-0.22%)	44.50 (-0.45%)	44.50 (-0.45%)	**21.20** (+23.98%)	**24.50** (+22.50%)	**24.60** (+23.00%)	**24.60** (+23.00%)
	+ImpartTail (Ours)	10	34.70 (-15.78%)	37.40 (-16.14%)	37.40 (-16.33%)	37.40 (-16.33%)	21.10 (+23.39%)	24.50 (+22.50%)	24.50 (+22.50%)	24.50 (+22.50%)
	SceneSayerODE [30]	-	**40.30**	**43.50**	**43.50**	**43.50**	17.50	20.70	20.90	20.90
	+ImpartTail (Ours)	70	38.00 (-5.71%)	41.10 (-5.52%)	41.10 (-5.52%)	41.10 (-5.52%)	**19.50** (+11.43%)	**23.70** (+14.49%)	**23.90** (+14.35%)	**23.90** (+14.35%)
	+ImpartTail (Ours)	40	39.00 (-3.23%)	42.30 (-2.76%)	42.30 (-2.76%)	42.30 (-2.76%)	15.00 (-14.29%)	18.00 (-13.04%)	18.10 (-13.40%)	18.10 (-13.40%)
	+ImpartTail (Ours)	10	27.50 (-31.76%)	29.40 (-32.41%)	29.40 (-32.41%)	29.40 (-32.41%)	19.30 (+10.29%)	23.20 (+12.08%)	23.50 (+12.44%)	23.50 (+12.44%)
	SceneSayerSDE [30]	-	**43.70**	**47.40**	**47.40**	**47.40**	18.20	21.70	21.80	21.80
	+ImpartTail (Ours)	70	40.90 (-6.41%)	44.40 (-6.33%)	44.40 (-6.33%)	44.40 (-6.33%)	22.40 (+23.08%)	27.00 (+24.42%)	27.30 (+25.23%)	27.30 (+25.23%)
	+ImpartTail (Ours)	40	43.20 (-1.14%)	46.90 (-1.05%)	46.90 (-1.05%)	46.90 (-1.05%)	17.80 (-2.20%)	21.20 (-2.30%)	21.40 (-1.83%)	21.40 (-1.83%)
	+ImpartTail (Ours)	10	35.30 (-19.22%)	37.90 (-20.04%)	37.90 (-20.04%)	37.90 (-20.04%)	**27.60** (+51.65%)	**33.40** (+53.92%)	**33.60** (+54.13%)	**33.60** (+54.13%)

CVPR
#16783

Table 18. **GAGS**-With Constraint-0.7 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.7	STTran++ [30]	-	37.30	40.50	40.50	40.50	21.90	25.00	25.00	25.00
	+ImparTail (Ours)	70	**49.50** (+32.71%)	**52.70** (+30.12%)	**52.70** (+30.12%)	**52.70** (+30.12%)	**26.20** (+19.63%)	**30.10** (+20.40%)	**30.10** (+20.40%)	**30.10** (+20.40%)
	+ImparTail (Ours)	40	44.80 (+20.11%)	47.70 (+17.78%)	47.70 (+17.78%)	47.70 (+17.78%)	24.00 (+9.59%)	26.90 (+7.60%)	26.90 (+7.60%)	26.90 (+7.60%)
	+ImparTail (Ours)	10	43.70 (+17.16%)	46.20 (+14.07%)	46.20 (+14.07%)	46.20 (+14.07%)	25.80 (+17.81%)	29.10 (+16.40%)	29.10 (+16.40%)	29.10 (+16.40%)
	DSGDetr++ [30]	-	48.50	51.80	51.80	51.80	20.80	23.80	23.80	23.80
	+ImparTail (Ours)	70	**49.80** (+2.68%)	**52.90** (+2.12%)	**53.00** (+2.32%)	**53.00** (+2.32%)	25.30 (+21.63%)	28.70 (+20.59%)	28.70 (+20.59%)	28.70 (+20.59%)
	+ImparTail (Ours)	40	49.20 (+1.44%)	52.30 (+0.97%)	52.30 (+0.97%)	52.30 (+0.97%)	26.90 (+29.33%)	30.70 (+28.99%)	30.80 (+29.41%)	30.80 (+29.41%)
	+ImparTail (Ours)	10	40.70 (-16.08%)	43.10 (-16.80%)	43.10 (-16.80%)	43.10 (-16.80%)	**28.30** (+36.06%)	**32.50** (+36.55%)	**32.50** (+36.55%)	**32.50** (+36.55%)
	SceneSayerODE [30]	-	**48.50**	**51.50**	**51.50**	**51.50**	20.70	24.00	24.00	24.00
	+ImparTail (Ours)	70	45.50 (-6.19%)	48.50 (-5.83%)	48.50 (-5.83%)	48.50 (-5.83%)	**23.50** (+13.53%)	**27.80** (+15.83%)	**27.90** (+16.25%)	**27.90** (+16.25%)
	+ImparTail (Ours)	40	47.50 (-2.06%)	50.60 (-1.75%)	50.60 (-1.75%)	50.60 (-1.75%)	18.10 (-12.56%)	21.30 (-11.25%)	21.30 (-11.25%)	21.30 (-11.25%)
	+ImparTail (Ours)	10	34.30 (-29.28%)	36.20 (-29.71%)	36.20 (-29.71%)	36.20 (-29.71%)	23.20 (+12.08%)	27.50 (+14.58%)	27.50 (+14.58%)	27.50 (+14.58%)
	SceneSayerSDE [30]	-	**50.90**	**54.10**	**54.10**	**54.10**	21.00	24.60	24.60	24.60
	+ImparTail (Ours)	70	48.50 (-4.72%)	51.70 (-4.44%)	51.70 (-4.44%)	51.70 (-4.44%)	26.50 (+26.19%)	31.40 (+27.64%)	31.50 (+28.05%)	31.50 (+28.05%)
	+ImparTail (Ours)	40	50.50 (-0.79%)	53.80 (-0.55%)	53.80 (-0.55%)	53.80 (-0.55%)	20.90 (-0.48%)	24.40 (-0.81%)	24.40 (-0.81%)	24.40 (-0.81%)
	+ImparTail (Ours)	10	41.50 (-18.47%)	43.90 (-18.85%)	43.90 (-18.85%)	43.90 (-18.85%)	**32.10** (+52.86%)	**38.40** (+56.10%)	**38.50** (+56.50%)	**38.50** (+56.50%)

Table 19. **GAGS**-With Constraint-0.9 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.9	STTran++ [30]	-	45.50	47.80	47.80	47.80	28.80	31.30	31.30	31.30
	+ImparTail (Ours)	70	**59.00** (+29.67%)	**61.30** (+28.24%)	**61.30** (+28.24%)	**61.30** (+28.24%)	34.20 (+18.75%)	37.20 (+18.85%)	37.30 (+19.17%)	37.30 (+19.17%)
	+ImparTail (Ours)	40	53.90 (+18.46%)	56.00 (+17.15%)	56.00 (+17.15%)	56.00 (+17.15%)	31.90 (+10.76%)	34.40 (+9.90%)	34.40 (+9.90%)	34.40 (+9.90%)
	+ImparTail (Ours)	10	50.90 (+11.87%)	52.70 (+10.25%)	52.70 (+10.25%)	52.70 (+10.25%)	**36.70** (+27.43%)	**40.30** (+28.75%)	**40.30** (+28.75%)	**40.30** (+28.75%)
	DSGDetr++ [30]	-	57.70	60.00	60.00	60.00	27.00	29.40	29.40	29.40
	+ImparTail (Ours)	70	**59.20** (+2.60%)	**61.40** (+2.33%)	**61.40** (+2.33%)	**61.40** (+2.33%)	34.20 (+26.67%)	36.90 (+25.51%)	36.90 (+25.51%)	36.90 (+25.51%)
	+ImparTail (Ours)	40	57.00 (-1.21%)	59.10 (-1.50%)	59.10 (-1.50%)	59.10 (-1.50%)	37.00 (+37.04%)	40.10 (+36.39%)	40.10 (+36.39%)	40.10 (+36.39%)
	+ImparTail (Ours)	10	46.10 (-20.10%)	47.60 (-20.67%)	47.60 (-20.67%)	47.60 (-20.67%)	**38.70** (+43.33%)	**42.80** (+45.58%)	**42.80** (+45.58%)	**42.80** (+45.58%)
	SceneSayerODE [30]	-	**58.50**	**60.60**	**60.60**	**60.60**	25.10	27.60	27.60	27.60
	+ImparTail (Ours)	70	54.10 (-7.52%)	56.10 (-7.43%)	56.10 (-7.43%)	56.10 (-7.43%)	27.70 (+10.36%)	30.60 (+10.87%)	30.60 (+10.87%)	30.60 (+10.87%)
	+ImparTail (Ours)	40	56.70 (-3.08%)	59.00 (-2.64%)	59.00 (-2.64%)	59.00 (-2.64%)	21.80 (-13.15%)	24.40 (-11.59%)	24.40 (-11.59%)	24.40 (-11.59%)
	+ImparTail (Ours)	10	43.80 (-25.13%)	45.30 (-25.25%)	45.30 (-25.25%)	45.30 (-25.25%)	**28.50** (+13.55%)	**31.60** (+14.49%)	**31.60** (+14.49%)	**31.60** (+14.49%)
	SceneSayerSDE [30]	-	**59.00**	**61.20**	**61.20**	**61.20**	24.70	27.30	27.30	27.30
	+ImparTail (Ours)	70	57.00 (-3.39%)	59.20 (-3.27%)	59.20 (-3.27%)	59.20 (-3.27%)	31.80 (+28.74%)	35.40 (+29.67%)	35.40 (+29.67%)	35.40 (+29.67%)
	+ImparTail (Ours)	40	58.10 (-1.53%)	60.30 (-1.47%)	60.30 (-1.47%)	60.30 (-1.47%)	24.60 (-0.40%)	27.30	27.30	27.30
	+ImparTail (Ours)	10	49.40 (-16.27%)	51.10 (-16.50%)	51.10 (-16.50%)	51.10 (-16.50%)	**38.60** (+56.28%)	**43.50** (+59.34%)	**43.50** (+59.34%)	**43.50** (+59.34%)

6.2.6. Results - With Constraint Setting - Partially Grounded Action Genome Scenes (PGAGS)

Table 20. **PGAGS**-With Constraint-0.3 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.3	STTran++ [30]	-	20.30	21.40	21.40	21.40	12.60	13.60	13.60	13.60
	+ImparTail (Ours)	70	26.70 (+31.53%)	27.70 (+29.44%)	27.70 (+29.44%)	27.70 (+29.44%)	12.80 (+1.59%)	13.60	13.60	13.60
	+ImparTail (Ours)	40	**27.60** (+35.96%)	**28.70** (+34.11%)	**28.70** (+34.11%)	**28.70** (+34.11%)	15.10 (+19.84%)	16.30 (+19.85%)	16.30 (+19.85%)	16.30 (+19.85%)
	+ImparTail (Ours)	10	21.80 (+7.39%)	22.40 (+4.67%)	22.40 (+4.67%)	22.40 (+4.67%)	**17.60** (+39.68%)	**18.90** (+38.97%)	**18.90** (+38.97%)	**18.90** (+38.97%)
	DSGDetr++ [30]	-	**28.10**	**29.30**	**29.30**	**29.30**	13.00	14.00	14.00	14.00
	+ImparTail (Ours)	70	26.50 (-5.69%)	27.60 (-5.80%)	27.60 (-5.80%)	27.60 (-5.80%)	14.70 (+13.08%)	15.80 (+12.86%)	15.80 (+12.86%)	15.80 (+12.86%)
	+ImparTail (Ours)	40	24.60 (-12.46%)	25.80 (-11.95%)	25.80 (-11.95%)	25.80 (-11.95%)	13.90 (+6.92%)	14.90 (+6.43%)	14.90 (+6.43%)	14.90 (+6.43%)
	+ImparTail (Ours)	10	25.30 (-9.96%)	26.30 (-10.24%)	26.30 (-10.24%)	26.30 (-10.24%)	**16.40** (+26.15%)	**17.80** (+27.14%)	**17.80** (+27.14%)	**17.80** (+27.14%)
	SceneSayerODE [30]	-	22.60	23.70	23.80	23.80	9.40	10.50	10.50	10.50
	+ImparTail (Ours)	70	18.10 (-19.91%)	18.80 (-20.68%)	18.80 (-21.01%)	18.80 (-21.01%)	10.50 (+11.70%)	11.80 (+12.38%)	11.80 (+12.38%)	11.80 (+12.38%)
	+ImparTail (Ours)	40	**28.00** (+23.89%)	**29.40** (+24.05%)	**29.40** (+23.53%)	**29.40** (+23.53%)	12.70 (+35.11%)	14.10 (+34.29%)	14.10 (+34.29%)	14.10 (+34.29%)
	+ImparTail (Ours)	10	21.60 (-4.42%)	22.30 (-5.91%)	22.40 (-5.88%)	22.40 (-5.88%)	**15.60** (+65.96%)	**17.80** (+69.52%)	**17.80** (+69.52%)	**17.80** (+69.52%)
	SceneSayerSDE [30]	-	**29.30**	**30.80**	**30.80**	**30.80**	13.50	15.20	15.20	15.20
	+ImparTail (Ours)	70	27.70 (-5.46%)	29.20 (-5.19%)	29.20 (-5.19%)	29.20 (-5.19%)	16.50 (+22.22%)	18.50 (+21.71%)	18.50 (+21.71%)	18.50 (+21.71%)
	+ImparTail (Ours)	40	28.60 (-2.39%)	30.10 (-2.27%)	30.10 (-2.27%)	30.10 (-2.27%)	14.10 (+4.44%)	15.80 (+3.95%)	15.80 (+3.95%)	15.80 (+3.95%)
	+ImparTail (Ours)	10	25.60 (-12.63%)	26.80 (-12.99%)	26.80 (-12.99%)	26.80 (-12.99%)	**20.30** (+50.37%)	**23.90** (+57.24%)	**23.90** (+57.24%)	**23.90** (+57.24%)

Table 21. **PGAGS**-With Constraint-0.5 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.5	STTran++ [30]	-	22.50	23.80	23.80	23.80	14.30	15.80	15.80	15.80
	+ImparTail (Ours)	70	29.50 (+31.11%)	30.60 (+28.57%)	30.60 (+28.57%)	30.60 (+28.57%)	14.50 (+1.40%)	15.70 (-0.63%)	15.70 (-0.63%)	15.70 (-0.63%)
	+ImparTail (Ours)	40	**30.20** (+34.22%)	**31.40** (+31.93%)	**31.40** (+31.93%)	**31.40** (+31.93%)	16.90 (+18.18%)	18.50 (+17.09%)	18.50 (+17.09%)	18.50 (+17.09%)
	+ImparTail (Ours)	10	24.10 (+7.11%)	25.00 (+5.04%)	25.00 (+5.04%)	25.00 (+5.04%)	**20.20** (+41.26%)	**22.00** (+39.24%)	**22.00** (+39.24%)	**22.00** (+39.24%)
	DSGDetr++ [30]	-	**31.70**	**33.00**	**33.00**	**33.00**	15.00	16.30	16.30	16.30
	+ImparTail (Ours)	70	28.70 (-9.46%)	29.90 (-9.39%)	29.90 (-9.39%)	29.90 (-9.39%)	16.80 (+12.00%)	18.30 (+12.27%)	18.30 (+12.27%)	18.30 (+12.27%)
	+ImparTail (Ours)	40	28.30 (-10.73%)	29.50 (-10.61%)	29.50 (-10.61%)	29.50 (-10.61%)	16.00 (+6.67%)	17.40 (+6.75%)	17.40 (+6.75%)	17.40 (+6.75%)
	+ImparTail (Ours)	10	28.50 (-10.09%)	29.50 (-10.61%)	29.50 (-10.61%)	29.50 (-10.61%)	**18.60** (+24.00%)	**20.10** (+23.31%)	**20.10** (+23.31%)	**20.10** (+23.31%)
	SceneSayerODE [30]	-	25.80	27.20	27.20	27.20	11.20	12.80	12.80	12.80
	+ImparTail (Ours)	70	21.20 (-17.83%)	22.20 (-18.38%)	22.20 (-18.38%)	22.20 (-18.38%)	12.60 (+12.50%)	14.50 (+13.28%)	14.50 (+13.28%)	14.50 (+13.28%)
	+ImparTail (Ours)	40	**31.90** (+23.64%)	**33.60** (+23.53%)	**33.60** (+23.53%)	**33.60** (+23.53%)	14.50 (+29.46%)	16.40 (+28.12%)	16.40 (+28.12%)	16.40 (+28.12%)
	+ImparTail (Ours)	10	25.30 (-1.94%)	26.30 (-3.31%)	26.30 (-3.31%)	26.30 (-3.31%)	**18.00** (+60.71%)	**21.30** (+66.41%)	**21.40** (+67.19%)	**21.40** (+67.19%)
	SceneSayerSDE [30]	-	**33.00**	**34.80**	**34.80**	**34.80**	15.20	17.50	17.50	17.50
	+ImparTail (Ours)	70	31.20 (-5.45%)	32.80 (-5.75%)	32.80 (-5.75%)	32.80 (-5.75%)	18.40 (+21.05%)	21.00 (+20.00%)	21.10 (+20.57%)	21.10 (+20.57%)
	+ImparTail (Ours)	40	32.30 (-2.12%)	34.10 (-2.01%)	34.10 (-2.01%)	34.10 (-2.01%)	15.70 (+3.29%)	17.90 (+2.29%)	17.90 (+2.29%)	17.90 (+2.29%)
	+ImparTail (Ours)	10	28.90 (-12.42%)	30.30 (-12.93%)	30.30 (-12.93%)	30.30 (-12.93%)	**22.20** (+46.05%)	**26.50** (+51.43%)	**26.50** (+51.43%)	**26.50** (+51.43%)

Table 22. **PGAGS**-With Constraint-0.7 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.7	STTran++ [30]	-	27.10	28.20	28.20	28.20	17.20	18.60	18.60	18.60
	+ImparTail (Ours)	70	**36.10** (+33.21%)	**37.10** (+31.56%)	**37.10** (+31.56%)	**37.10** (+31.56%)	18.70 (+8.72%)	19.90 (+6.99%)	19.90 (+6.99%)	19.90 (+6.99%)
	+ImparTail (Ours)	40	36.00 (+32.84%)	37.00 (+31.21%)	37.00 (+31.21%)	37.00 (+31.21%)	21.00 (+22.09%)	22.70 (+22.04%)	22.70 (+22.04%)	22.70 (+22.04%)
	+ImparTail (Ours)	10	29.60 (+9.23%)	30.20 (+7.09%)	30.20 (+7.09%)	30.20 (+7.09%)	**25.60** (+48.84%)	**27.40** (+47.31%)	**27.40** (+47.31%)	**27.40** (+47.31%)
	DSGDetr++ [30]	-	**37.30**	**38.50**	**38.50**	**38.50**	18.10	19.40	19.40	19.40
	+ImparTail (Ours)	70	34.30 (-8.04%)	35.40 (-8.05%)	35.40 (-8.05%)	35.40 (-8.05%)	20.70 (+14.36%)	22.10 (+13.92%)	22.10 (+13.92%)	22.10 (+13.92%)
	+ImparTail (Ours)	40	33.90 (-9.12%)	34.90 (-9.35%)	34.90 (-9.35%)	34.90 (-9.35%)	20.60 (+13.81%)	21.80 (+12.37%)	21.80 (+12.37%)	21.80 (+12.37%)
	+ImparTail (Ours)	10	33.40 (-10.46%)	34.50 (-10.39%)	34.50 (-10.39%)	34.50 (-10.39%)	**23.60** (+30.39%)	**25.50** (+31.44%)	**25.50** (+31.44%)	**25.50** (+31.44%)
	SceneSayerODE [30]	-	30.80	32.20	32.20	32.20	13.60	15.10	15.10	15.10
	+ImparTail (Ours)	70	25.60 (-16.88%)	26.60 (-17.39%)	26.60 (-17.39%)	26.60 (-17.39%)	14.70 (+8.09%)	16.70 (+10.60%)	16.70 (+10.60%)	16.70 (+10.60%)
	+ImparTail (Ours)	40	**37.20** (+20.78%)	**38.70** (+20.19%)	**38.70** (+20.19%)	**38.70** (+20.19%)	16.90 (+24.26%)	18.70 (+23.84%)	18.80 (+24.50%)	18.80 (+24.50%)
	+ImparTail (Ours)	10	31.10 (+0.97%)	32.10 (-0.31%)	32.10 (-0.31%)	32.10 (-0.31%)	**22.20** (+63.24%)	**25.60** (+69.54%)	**25.70** (+70.20%)	**25.70** (+70.20%)
	SceneSayerSDE [30]	-	**38.80**	**40.30**	**40.30**	**40.30**	17.90	19.90	19.90	19.90
	+ImparTail (Ours)	70	36.80 (-5.15%)	38.30 (-4.96%)	38.30 (-4.96%)	38.30 (-4.96%)	21.70 (+21.23%)	24.00 (+20.60%)	24.00 (+20.60%)	24.00 (+20.60%)
	+ImparTail (Ours)	40	38.40 (-1.03%)	40.00 (-0.74%)	40.00 (-0.74%)	40.00 (-0.74%)	18.80 (+5.03%)	20.70 (+4.02%)	20.70 (+4.02%)	20.70 (+4.02%)
	+ImparTail (Ours)	10	34.40 (-11.34%)	35.60 (-11.66%)	35.60 (-11.66%)	35.60 (-11.66%)	**25.90** (+44.69%)	**30.00** (+50.75%)	**30.10** (+51.26%)	**30.10** (+51.26%)

Table 23. **PGAGS**-With Constraint-0.9 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.9	STTran++ [30]	-	31.00	31.80	31.80	31.80	21.20	22.50	22.50	22.50
	+ImparTail (Ours)	70	**42.00** (+35.48%)	**42.80** (+34.59%)	**42.80** (+34.59%)	**42.80** (+34.59%)	26.00 (+22.64%)	27.20 (+20.89%)	27.20 (+20.89%)	27.20 (+20.89%)
	+ImparTail (Ours)	40	39.90 (+28.71%)	40.60 (+27.67%)	40.60 (+27.67%)	40.60 (+27.67%)	26.60 (+25.47%)	28.10 (+24.89%)	28.10 (+24.89%)	28.10 (+24.89%)
	+ImparTail (Ours)	10	32.70 (+5.48%)	33.20 (+4.40%)	33.20 (+4.40%)	33.20 (+4.40%)	**32.60** (+53.77%)	**34.20** (+52.00%)	**34.20** (+52.00%)	**34.20** (+52.00%)
	DSGDetr++ [30]	-	**43.10**	**44.00**	**44.00**	**44.00**	22.20	23.40	23.40	23.40
	+ImparTail (Ours)	70	40.00 (-7.19%)	40.80 (-7.27%)	40.80 (-7.27%)	40.80 (-7.27%)	24.90 (+12.16%)	26.20 (+11.97%)	26.20 (+11.97%)	26.20 (+11.97%)
	+ImparTail (Ours)	40	39.60 (-8.12%)	40.50 (-7.95%)	40.50 (-7.95%)	40.50 (-7.95%)	26.20 (+18.02%)	27.50 (+17.52%)	27.50 (+17.52%)	27.50 (+17.52%)
	+ImparTail (Ours)	10	38.50 (-10.67%)	39.20 (-10.91%)	39.20 (-10.91%)	39.20 (-10.91%)	**30.90** (+39.19%)	**32.20** (+37.61%)	**32.20** (+37.61%)	**32.20** (+37.61%)
	SceneSayerODE [30]	-	36.60	37.60	37.60	37.60	16.60	17.90	17.90	17.90
	+ImparTail (Ours)	70	31.80 (-13.11%)	32.60 (-13.30%)	32.60 (-13.30%)	32.60 (-13.30%)	18.40 (+10.84%)	20.10 (+12.29%)	20.10 (+12.29%)	20.10 (+12.29%)
	+ImparTail (Ours)	40	**42.70** (+16.67%)	**43.80** (+16.49%)	**43.80** (+16.49%)	**43.80** (+16.49%)	19.40 (+16.87%)	20.90 (+16.76%)	20.90 (+16.76%)	20.90 (+16.76%)
	+ImparTail (Ours)	10	38.40 (+4.92%)	39.30 (+4.52%)	39.30 (+4.52%)	39.30 (+4.52%)	**28.30** (+70.48%)	**30.90** (+72.63%)	**30.90** (+72.63%)	**30.90** (+72.63%)
	SceneSayerSDE [30]	-	44.40	45.50	45.50	45.50	21.00	22.60	22.60	22.60
	+ImparTail (Ours)	70	43.10 (-2.93%)	44.20 (-2.86%)	44.20 (-2.86%)	44.20 (-2.86%)	25.20 (+20.00%)	27.00 (+19.47%)	27.00 (+19.47%)	27.00 (+19.47%)
	+ImparTail (Ours)	40	**44.70** (+0.68%)	**45.90** (+0.88%)	**45.90** (+0.88%)	**45.90** (+0.88%)	21.80 (+3.81%)	23.50 (+3.98%)	23.50 (+3.98%)	23.50 (+3.98%)
	+ImparTail (Ours)	10	40.50 (-8.78%)	41.40 (-9.01%)	41.40 (-9.01%)	41.40 (-9.01%)	**31.60** (+50.48%)	**34.70** (+53.54%)	**34.70** (+53.54%)	**34.70** (+53.54%)

6.2.7. Results - With Constraint Setting - Action Genome Scenes (AGS)

Table 24. **AGS**-With Constraint-0.3 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.3	STTran++ [30]	-	19.40	**31.10**	**33.50**	**33.50**	7.60	15.30	16.90	16.90
	+ImparTail (Ours)	70	**19.70** (+1.55%)	30.20 (-2.89%)	32.40 (-3.28%)	32.40 (-3.28%)	7.80 (+2.63%)	15.10 (-1.31%)	16.80 (-0.59%)	16.80 (-0.59%)
	+ImparTail (Ours)	40	19.20 (-1.03%)	30.30 (-2.57%)	32.60 (-2.69%)	32.60 (-2.69%)	8.40 (+10.53%)	16.80 (+9.80%)	18.60 (+10.06%)	18.60 (+10.06%)
	+ImparTail (Ours)	10	18.30 (-5.67%)	27.40 (-11.90%)	29.30 (-12.54%)	29.30 (-12.54%)	**8.90** (+17.11%)	**17.90** (+16.99%)	**19.70** (+16.57%)	**19.70** (+16.57%)
	DSGDetr++ [30]	-	**18.80**	**28.30**	**29.90**	**29.90**	7.10	12.80	13.80	13.80
	+ImparTail (Ours)	70	18.60 (-1.06%)	27.40 (-3.18%)	29.10 (-2.68%)	29.10 (-2.68%)	8.20 (+15.49%)	15.20 (+18.75%)	16.60 (+20.29%)	16.60 (+20.29%)
	+ImparTail (Ours)	40	16.50 (-12.23%)	24.70 (-12.72%)	26.30 (-12.04%)	26.30 (-12.04%)	8.30 (+16.90%)	15.80 (+23.44%)	17.10 (+23.91%)	17.10 (+23.91%)
	+ImparTail (Ours)	10	16.90 (-10.11%)	23.50 (-16.96%)	24.50 (-18.06%)	24.50 (-18.06%)	**9.30** (+30.99%)	**18.40** (+43.75%)	**20.20** (+46.38%)	**20.20** (+46.38%)
	SceneSayerODE [30]	-	15.10	25.70	30.10	30.20	5.50	11.90	19.00	19.00
	+ImparTail (Ours)	70	13.80 (-8.61%)	23.00 (-10.51%)	26.80 (-10.96%)	26.80 (-11.26%)	5.80 (+5.45%)	12.80 (+7.56%)	16.40 (-13.68%)	16.40 (-13.68%)
	+ImparTail (Ours)	40	**16.30** (+7.95%)	**28.20** (+9.73%)	**33.70** (+11.96%)	**33.80** (+11.92%)	5.80 (+5.45%)	12.50 (+5.04%)	16.00 (-15.79%)	16.10 (-15.26%)
	+ImparTail (Ours)	10	14.80 (-1.99%)	24.80 (-3.50%)	29.10 (-3.32%)	29.10 (-3.64%)	**6.60** (+20.00%)	**15.40** (+29.41%)	**20.40** (+7.37%)	**20.40** (+7.37%)
	SceneSayerSDE [30]	-	15.90	27.80	33.40	33.50	5.90	13.40	17.20	17.20
	+ImparTail (Ours)	70	**16.10** (+1.26%)	28.10 (+1.08%)	33.70 (+0.90%)	33.70 (+0.60%)	6.70 (+13.56%)	15.80 (+17.91%)	20.80 (+20.93%)	20.80 (+20.93%)
	+ImparTail (Ours)	40	16.10 (+1.26%)	**28.20** (+1.44%)	**33.90** (+1.50%)	**34.00** (+1.49%)	5.90	13.10 (-2.24%)	17.20	17.20
	+ImparTail (Ours)	10	13.10 (-17.61%)	23.00 (-17.27%)	27.20 (-18.56%)	27.20 (-18.81%)	**7.30** (+23.73%)	**19.30** (+44.03%)	**25.90** (+50.58%)	**26.00** (+51.16%)

Table 25. **AGS**-With Constraint-0.5 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.5	STTran++ [30]	-	20.40	**33.30**	**35.90**	**35.90**	7.90	16.40	18.40	18.40
	+ImparTail (Ours)	70	**21.10** (+3.43%)	32.70 (-1.80%)	34.90 (-2.79%)	34.90 (-2.79%)	8.30 (+5.06%)	16.50 (+0.61%)	18.30 (-0.54%)	18.30 (-0.54%)
	+ImparTail (Ours)	40	20.60 (+0.98%)	32.50 (-2.40%)	35.00 (-2.51%)	35.00 (-2.51%)	8.90 (+12.66%)	17.80 (+8.54%)	19.80 (+7.61%)	19.80 (+7.61%)
	+ImparTail (Ours)	10	19.80 (-2.94%)	29.70 (-10.81%)	31.80 (-11.42%)	31.80 (-11.42%)	**9.30** (+17.72%)	**18.70** (+14.02%)	**20.90** (+13.59%)	**20.90** (+13.59%)
	DSGDetr++ [30]	-	19.80	**30.00**	**31.90**	**31.90**	7.40	13.40	14.60	14.60
	+ImparTail (Ours)	70	**20.10** (+1.52%)	29.60 (-1.33%)	31.50 (-1.25%)	31.50 (-1.25%)	8.70 (+17.57%)	16.40 (+22.39%)	17.90 (+22.60%)	17.90 (+22.60%)
	+ImparTail (Ours)	40	18.10 (-8.59%)	27.00 (-10.00%)	28.80 (-9.72%)	28.80 (-9.72%)	8.90 (+20.27%)	17.00 (+26.87%)	18.60 (+27.40%)	18.60 (+27.40%)
	+ImparTail (Ours)	10	18.10 (-8.59%)	25.30 (-15.67%)	26.70 (-16.30%)	26.70 (-16.30%)	**9.80** (+32.43%)	**19.80** (+47.76%)	**21.90** (+50.00%)	**21.90** (+50.00%)
	SceneSayerODE [30]	-	16.60	28.20	33.50	33.60	5.80	12.60	16.90	16.90
	+ImparTail (Ours)	70	15.10 (-9.04%)	25.40 (-9.93%)	30.00 (-10.45%)	30.00 (-10.71%)	5.90 (+1.72%)	13.50 (+7.14%)	18.00 (+6.51%)	18.00 (+6.51%)
	+ImparTail (Ours)	40	**17.30** (+4.22%)	**29.90** (+6.03%)	**36.20** (+8.06%)	**36.30** (+8.04%)	6.00 (+3.45%)	12.70 (+0.79%)	16.90	16.90
	+ImparTail (Ours)	10	15.60 (-6.02%)	26.20 (-7.09%)	31.30 (-6.57%)	31.40 (-6.55%)	**6.80** (+17.24%)	**16.10** (+27.78%)	**22.00** (+30.18%)	**22.00** (+30.18%)
	SceneSayerSDE [30]	-	**17.50**	30.00	36.50	36.50	6.40	13.70	18.30	18.30
	+ImparTail (Ours)	70	16.90 (-3.43%)	29.70 (-1.00%)	36.20 (-0.82%)	36.20 (-0.82%)	6.80 (+6.25%)	16.10 (+17.52%)	21.90 (+19.67%)	21.90 (+19.67%)
	+ImparTail (Ours)	40	17.40 (-0.57%)	**30.20** (+0.67%)	**36.70** (+0.55%)	**36.70** (+0.55%)	6.30 (-1.56%)	13.70	18.40 (+0.55%)	18.40 (+0.55%)
	+ImparTail (Ours)	10	14.50 (-17.14%)	24.80 (-17.33%)	29.90 (-18.08%)	30.00 (-17.81%)	**7.40** (+15.62%)	**19.10** (+39.42%)	**27.70** (+51.37%)	**27.80** (+51.91%)

Table 26. **AGS**-With Constraint-0.7 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.7	STTran++ [30]	-	22.50	36.40	**39.20**	**39.20**	9.10	18.20	20.20	20.20
	+ImparTail (Ours)	70	**23.50** (+4.44%)	**36.50** (+0.27%)	39.10 (-0.26%)	39.10 (-0.26%)	9.90 (+8.79%)	19.30 (+6.04%)	21.50 (+6.44%)	21.50 (+6.44%)
	+ImparTail (Ours)	40	23.00 (+2.22%)	36.00 (-1.10%)	38.70 (-1.28%)	38.70 (-1.28%)	10.40 (+14.29%)	20.50 (+12.64%)	22.80 (+12.87%)	22.80 (+12.87%)
	+ImparTail (Ours)	10	21.90 (-2.67%)	32.70 (-10.16%)	34.90 (-10.97%)	34.90 (-10.97%)	**10.90** (+19.78%)	**21.90** (+20.33%)	**24.10** (+19.31%)	**24.10** (+19.31%)
	DSGDetr++ [30]	-	22.20	**33.20**	**35.10**	**35.10**	8.40	14.80	16.00	16.00
	+ImparTail (Ours)	70	**22.40** (+0.90%)	32.60 (-1.81%)	34.50 (-1.71%)	34.50 (-1.71%)	10.10 (+20.24%)	18.40 (+24.32%)	20.10 (+25.63%)	20.10 (+25.63%)
	+ImparTail (Ours)	40	20.50 (-7.66%)	30.60 (-7.83%)	32.60 (-7.12%)	32.60 (-7.12%)	10.50 (+25.00%)	19.50 (+31.76%)	21.20 (+32.50%)	21.20 (+32.50%)
	+ImparTail (Ours)	10	20.20 (-9.01%)	28.10 (-15.36%)	29.60 (-15.67%)	29.60 (-15.67%)	**11.50** (+36.90%)	**23.10** (+56.08%)	**25.20** (+57.50%)	**25.20** (+57.50%)
	SceneSayerODE [30]	-	19.00	32.00	37.90	38.00	6.70	14.00	18.50	18.50
	+ImparTail (Ours)	70	17.60 (-7.37%)	28.90 (-9.69%)	33.90 (-10.55%)	33.90 (-10.79%)	6.90 (+2.99%)	15.00 (+7.14%)	19.60 (+5.95%)	19.60 (+5.95%)
	+ImparTail (Ours)	40	**19.50** (+2.63%)	**33.10** (+3.44%)	**39.50** (+4.22%)	**39.50** (+3.95%)	6.80 (+1.49%)	13.90 (-0.71%)	18.20 (-1.62%)	18.30 (-1.08%)
	+ImparTail (Ours)	10	17.50 (-7.89%)	28.90 (-9.69%)	34.10 (-10.03%)	34.20 (-10.00%)	**7.90** (+17.91%)	**17.70** (+26.43%)	**23.40** (+26.49%)	**23.40** (+26.49%)
	SceneSayerSDE [30]	-	19.50	33.00	39.60	39.70	7.10	14.60	19.30	19.30
	+ImparTail (Ours)	70	19.10 (-2.05%)	32.50 (-1.52%)	39.40 (-0.51%)	39.40 (-0.76%)	7.80 (+9.86%)	17.40 (+19.18%)	23.80 (+23.32%)	23.80 (+23.32%)
	+ImparTail (Ours)	40	**19.60** (+0.51%)	**33.50** (+1.52%)	**40.40** (+2.02%)	**40.40** (+1.76%)	7.10	15.00 (+2.74%)	20.20 (+4.66%)	20.20 (+4.66%)
	+ImparTail (Ours)	10	16.80 (-13.85%)	27.90 (-15.45%)	33.00 (-16.67%)	33.00 (-16.88%)	**8.60** (+21.13%)	**21.30** (+45.89%)	**29.30** (+51.81%)	**29.30** (+51.81%)

Table 27. **AGS**-With Constraint-0.9 results for SGA.

$\mathcal{F}$	Method	$\mathcal{S}$	WITH CONSTRAINT							
			R@10	R@20	R@50	R@100	mR@10	mR@20	mR@50	mR@100
0.9	STTran++ [30]	-	24.60	39.80	42.90	42.90	9.80	20.90	23.50	23.50
	+ImparTail (Ours)	70	**25.80** (+4.88%)	**40.50** (+1.76%)	**43.50** (+1.40%)	**43.50** (+1.40%)	11.10 (+13.27%)	22.70 (+8.61%)	25.40 (+8.09%)	25.40 (+8.09%)
	+ImparTail (Ours)	40	24.70 (+0.41%)	39.50 (-0.75%)	42.50 (-0.93%)	42.50 (-0.93%)	12.30 (+25.51%)	25.50 (+22.01%)	28.20 (+20.00%)	28.20 (+20.00%)
	+ImparTail (Ours)	10	23.60 (-4.07%)	35.60 (-10.55%)	37.90 (-11.66%)	37.90 (-11.66%)	**12.60** (+28.57%)	**27.10** (+29.67%)	**29.90** (+27.23%)	**29.90** (+27.23%)
	DSGDetr++ [30]	-	**24.80**	**37.50**	**39.70**	**39.70**	9.50	17.70	19.20	19.20
	+ImparTail (Ours)	70	24.70 (-0.40%)	37.10 (-1.07%)	39.50 (-0.50%)	39.50 (-0.50%)	11.90 (+25.26%)	22.50 (+27.12%)	24.90 (+29.69%)	24.90 (+29.69%)
	+ImparTail (Ours)	40	24.00 (-3.23%)	36.20 (-3.47%)	38.50 (-3.02%)	38.50 (-3.02%)	12.40 (+30.53%)	24.80 (+40.11%)	27.00 (+40.62%)	27.00 (+40.62%)
	+ImparTail (Ours)	10	23.10 (-6.85%)	32.30 (-13.87%)	33.90 (-14.61%)	33.90 (-14.61%)	**13.60** (+43.16%)	**27.80** (+57.06%)	**30.20** (+57.29%)	**30.20** (+57.29%)
	SceneSayerODE [30]	-	**21.00**	**36.00**	43.20	43.20	7.00	15.60	21.00	21.00
	+ImparTail (Ours)	70	19.50 (-7.14%)	32.80 (-8.89%)	38.70 (-10.42%)	38.80 (-10.19%)	7.50 (+7.14%)	16.60 (+6.41%)	21.80 (+3.81%)	21.90 (+4.29%)
	+ImparTail (Ours)	40	20.90 (-0.48%)	35.90 (-0.28%)	**43.50** (+0.69%)	**43.50** (+0.69%)	7.00	15.20 (-2.56%)	20.60 (-1.90%)	20.60 (-1.90%)
	+ImparTail (Ours)	10	18.40 (-12.38%)	31.50 (-12.50%)	37.30 (-13.66%)	37.30 (-13.66%)	**8.30** (+18.57%)	**19.80** (+26.92%)	**26.10** (+24.29%)	**26.10** (+24.29%)
	SceneSayerSDE [30]	-	21.00	36.40	43.90	44.00	7.40	16.00	21.10	21.10
	+ImparTail (Ours)	70	20.60 (-1.90%)	35.90 (-1.37%)	43.80 (-0.23%)	43.80 (-0.45%)	8.20 (+10.81%)	18.90 (+18.12%)	25.60 (+21.33%)	25.60 (+21.33%)
	+ImparTail (Ours)	40	**21.20** (+0.95%)	**36.60** (+0.55%)	**44.40** (+1.14%)	**44.50** (+1.14%)	7.20 (-2.70%)	16.40 (+2.50%)	22.10 (+4.74%)	22.10 (+4.74%)
	+ImparTail (Ours)	10	18.70 (-10.95%)	31.50 (-13.46%)	37.60 (-14.35%)	37.60 (-14.55%)	**9.40** (+27.03%)	**24.00** (+50.00%)	**32.70** (+54.98%)	**32.70** (+54.98%)

6.3. Robust Video Scene Graph Generation

6.3.1. Findings

Table 28, Table 29 present the Robustness Evaluation Results for SGCLS and PREDCLS for Scene Graph Generation (SGG) under various corruption scenarios. These experiments assess how well models, with and without **IMPARTAIL**, handle different levels of data corruption. The settings include 15 corruption types and three graph-building strategies (With Cosntraint, No Constraint, Semi Constraint). Results highlight the impact of **IMPARTAIL** in improving robustness across these scenarios. **IMPARTAIL** performs best against Fog, Brightness, Saturate and moderate gains under Defocus blur, Gaussian Blur. **IMPARTAIL** shows an average of 25% gains for With Constraint mR@50, 10% gains for No Constraint mR@50.

6.3.2. Results

Table 28. Robustness Evaluation Results for SGG.

Severity	Mode	Corruption	Method	With Constraint			No Constraint						Semi Constraint		
				mR@10	mR@20	mR@50	R@10	R@20	R@50	mR@10	mR@20	mR@50	mR@10	mR@20	mR@50
3	sgcls	Gaussian Noise	DSGDetr [9]	9.6	10.3	10.3	20.9	25.4	26.8	15.7	19.4	23.4	11.4	15.3	15.7
			+IMPARTAIL (Ours)	**13.7** (+42.7%)	**14.9** (+44.7%)	**15.0** (+45.6%)	**21.0** (+0.5%)	**27.3** (+7.5%)	**30.1** (+12.3%)	**20.8** (+32.5%)	**25.4** (+30.9%)	**29.1** (+24.4%)	**15.6** (+36.8%)	**21.5** (+40.5%)	**22.2** (+41.4%)
		Shot Noise	DSGDetr [9]	10.0	10.8	10.8	21.9	26.6	28.1	16.6	20.4	26.5	12.1	16.3	16.6
			+IMPARTAIL (Ours)	**15.0** (+50.0%)	**16.5** (+52.8%)	**16.5** (+52.8%)	**22.6** (+3.2%)	**29.4** (+10.5%)	**32.6** (+16.0%)	**22.4** (+34.9%)	**27.3** (+33.8%)	**31.0** (+17.0%)	**17.2** (+42.1%)	**23.3** (+42.9%)	**24.2** (+45.8%)
		Impulse Noise	DSGDetr [9]	8.7	9.4	9.5	19.2	23.4	24.7	14.4	17.7	23.4	10.4	14.2	14.6
			+IMPARTAIL (Ours)	**12.4** (+42.5%)	**13.7** (+45.7%)	**13.7** (+44.2%)	**19.4** (+1.0%)	**25.3** (+8.1%)	**27.9** (+13.0%)	**18.5** (+28.5%)	**22.8** (+28.8%)	**26.2** (+12.0%)	**13.9** (+33.7%)	**19.2** (+35.2%)	**20.1** (+37.7%)
		Speckle Noise	DSGDetr [9]	12.6	13.7	13.7	26.2	32.2	34.2	20.3	25.0	32.2	15.0	20.7	21.2
			+IMPARTAIL (Ours)	**17.6** (+39.7%)	**19.3** (+40.9%)	**19.3** (+40.9%)	**26.4** (+0.8%)	**34.5** (+7.1%)	**38.4** (+12.3%)	**27.1** (+33.5%)	**32.6** (+30.4%)	**37.3** (+15.8%)	**20.4** (+36.0%)	**27.8** (+34.3%)	**28.9** (+36.3%)
		Gaussian Blur	DSGDetr [9]	21.1	22.8	22.8	**43.7**	**53.3**	56.5	33.4	41.0	52.3	24.6	33.2	34.0
			+IMPARTAIL (Ours)	**26.1** (+23.7%)	**28.7** (+25.9%)	**28.8** (+26.3%)	39.4 (-9.8%)	52.0 (-2.4%)	**57.8** (+2.3%)	**39.5** (+18.3%)	**48.8** (+19.0%)	**58.2** (+11.3%)	**29.4** (+19.5%)	**40.8** (+22.9%)	**42.5** (+25.0%)
		Defocus Blur	DSGDetr [9]	20.9	22.6	22.6	**43.2**	**52.8**	55.9	32.7	40.4	51.7	24.3	32.7	33.5
			+IMPARTAIL (Ours)	**25.6** (+22.5%)	**28.1** (+24.3%)	**28.1** (+24.3%)	39.0 (-9.7%)	51.4 (-2.7%)	**57.3** (+2.5%)	**38.6** (+18.0%)	**47.7** (+18.1%)	**57.0** (+10.3%)	**28.7** (+18.1%)	**40.0** (+22.3%)	**41.7** (+24.5%)
		Fog	DSGDetr [9]	22.6	24.9	24.9	**47.8**	**58.4**	62.0	35.5	43.4	54.6	26.6	36.1	37.2
			+IMPARTAIL (Ours)	**28.3** (+25.2%)	**31.8** (+27.7%)	**31.9** (+28.1%)	42.6 (-10.9%)	56.0 (-4.1%)	**62.1** (+0.2%)	**43.8** (+23.4%)	**53.0** (+22.1%)	**61.7** (+13.0%)	**31.8** (+19.5%)	**45.7** (+26.6%)	**48.2** (+29.6%)
		Frost	DSGDetr [9]	16.7	18.5	18.5	**34.5**	**42.3**	45.1	26.8	33.0	40.1	19.6	26.8	27.7
			+IMPARTAIL (Ours)	**22.4** (+34.1%)	**25.0** (+35.1%)	**25.1** (+35.7%)	31.9 (-7.5%)	42.0 (-0.7%)	**47.1** (+4.4%)	**34.9** (+30.2%)	**42.3** (+28.2%)	**48.2** (+20.2%)	**25.9** (+32.1%)	**36.5** (+36.2%)	**38.4** (+38.6%)
		Spatter	DSGDetr [9]	18.6	20.3	20.3	**41.4**	**51.0**	54.3	29.0	36.5	48.0	21.7	29.3	30.1
			+IMPARTAIL (Ours)	**24.7** (+32.8%)	**27.4** (+35.0%)	**27.5** (+35.5%)	38.6 (-6.8%)	50.6 (-0.8%)	**56.1** (+3.3%)	**37.4** (+29.0%)	**46.1** (+26.3%)	**55.8** (+16.2%)	**27.9** (+28.6%)	**39.3** (+34.1%)	**41.4** (+37.5%)
		Contrast	DSGDetr [9]	20.0	21.8	21.8	**42.4**	**52.2**	**55.5**	31.3	38.6	48.8	23.6	31.9	32.9
			+IMPARTAIL (Ours)	**24.9** (+24.5%)	**27.7** (+27.1%)	**27.8** (+27.5%)	36.9 (-13.0%)	49.0 (-6.1%)	54.6 (-1.6%)	**37.9** (+21.1%)	**46.0** (+19.2%)	**54.3** (+11.3%)	**27.5** (+16.5%)	**38.6** (+21.0%)	**40.6** (+23.4%)
		Brightness	DSGDetr [9]	23.6	25.7	25.7	**50.8**	**61.9**	65.5	36.8	45.4	57.5	27.6	37.5	38.6
			+IMPARTAIL (Ours)	**29.8** (+26.3%)	**33.2** (+29.2%)	**33.2** (+29.2%)	45.5 (-10.4%)	59.9 (-3.2%)	**66.1** (+0.9%)	**45.0** (+22.3%)	**55.4** (+22.0%)	**65.3** (+13.6%)	**33.7** (+22.1%)	**48.1** (+28.3%)	**50.6** (+31.1%)
		Pixelate	DSGDetr [9]	21.6	23.3	23.3	**48.1**	**59.7**	63.8	33.5	42.6	56.9	25.4	33.7	34.6
			+IMPARTAIL (Ours)	**27.7** (+28.2%)	**30.5** (+30.9%)	**30.5** (+30.9%)	43.8 (-8.9%)	57.7 (-3.4%)	**64.1** (+0.5%)	**42.5** (+26.9%)	**52.8** (+23.9%)	**63.1** (+10.9%)	**31.4** (+23.6%)	**44.5** (+32.0%)	**46.7** (+35.0%)
		Compression	DSGDetr [9]	19.9	21.5	21.6	**45.0**	**55.7**	59.5	31.3	40.1	52.4	23.4	31.2	32.1
			+IMPARTAIL (Ours)	**26.6** (+33.7%)	**29.3** (+36.3%)	**29.4** (+36.1%)	41.9 (-6.9%)	55.1 (-1.1%)	**61.5** (+3.4%)	**40.8** (+30.4%)	**50.5** (+25.9%)	**60.2** (+14.9%)	**29.8** (+27.4%)	**42.1** (+34.9%)	**44.0** (+37.1%)
		Sun Glare	DSGDetr [9]	12.1	13.2	13.2	**26.3**	32.5	34.7	19.3	24.4	30.2	14.2	19.2	19.6
			+IMPARTAIL (Ours)	**17.3** (+43.0%)	**19.4** (+47.0%)	**19.4** (+47.0%)	25.8 (-1.9%)	**34.3** (+5.5%)	**38.5** (+11.0%)	**26.6** (+37.8%)	**32.2** (+32.0%)	**37.3** (+23.5%)	**19.4** (+36.6%)	**27.6** (+43.7%)	**29.0** (+48.0%)
		Dust	DSGDetr [9]	13.2	14.5	14.6	**28.5**	**35.4**	**37.7**	21.5	26.8	34.2	15.4	20.8	21.4
			+IMPARTAIL (Ours)	**16.6** (+25.8%)	**18.6** (+28.3%)	**18.6** (+27.4%)	25.1 (-11.9%)	32.9 (-7.1%)	36.8 (-2.4%)	**25.9** (+20.5%)	**31.0** (+15.7%)	**37.0** (+8.2%)	**19.1** (+24.0%)	**26.3** (+26.4%)	**27.4** (+28.0%)
		Saturate	DSGDetr [9]	25.9	28.4	28.4	**54.6**	**66.2**	69.9	40.7	49.3	62.4	30.6	41.7	42.9
			+IMPARTAIL (Ours)	**31.5** (+21.6%)	**35.1** (+23.6%)	**35.2** (+23.9%)	48.8 (-10.6%)	63.8 (-3.6%)	**70.4** (+0.7%)	**47.4** (+16.5%)	**58.4** (+18.5%)	**68.8** (+10.3%)	**35.7** (+16.7%)	**51.0** (+22.3%)	**53.5** (+24.7%)

CVPR
#16783

Table 29. Robustness Evaluation Results for SGG.

Severity	Mode	Corruption	Method	With Constraint			No Constraint						Semi Constraint		
				mR@10	mR@20	mR@50	R@10	R@20	R@50	mR@10	mR@20	mR@50	mR@10	mR@20	mR@50
5	predcls	Gaussian Noise	STTran [5]	20.0	22.3	22.4	**64.2**	**87.6**	99.0	31.4	52.5	79.7	26.0	36.6	38.5
			+ImpartTail (Ours)	**37.6** (+88.0%)	**43.8** (+96.4%)	**43.9** (+96.0%)	62.5 (-2.6%)	84.6 (-3.4%)	**99.0** (0.0%)	**57.5** (+83.1%)	**77.7** (+48.0%)	**92.7** (+16.3%)	**42.2** (+62.3%)	**60.0** (+63.9%)	**62.9** (+63.4%)
		Shot Noise	STTran [5]	20.3	22.8	22.9	**64.6**	**88.0**	99.0	32.1	53.4	79.9	26.4	37.1	39.0
			+ImpartTail (Ours)	**37.5** (+84.7%)	**43.7** (+91.7%)	**43.8** (+91.3%)	62.4 (-3.4%)	84.6 (-3.9%)	**99.0** (0.0%)	**57.1** (+77.9%)	**77.9** (+45.9%)	**92.8** (+16.1%)	**42.2** (+59.8%)	**60.5** (+63.1%)	**63.6** (+63.1%)
		Impulse Noise	STTran [5]	20.3	22.8	22.9	**64.6**	**88.0**	99.0	31.9	52.9	79.8	26.3	37.2	39.1
			+ImpartTail (Ours)	**37.7** (+85.7%)	**43.8** (+92.1%)	**43.9** (+91.7%)	62.4 (-3.4%)	84.7 (-3.8%)	**99.0** (0.0%)	**57.4** (+79.9%)	**77.7** (+46.9%)	**92.6** (+16.0%)	**42.0** (+59.7%)	**60.1** (+61.6%)	**63.1** (+61.4%)
		Speckle Noise	STTran [5]	23.1	26.1	26.2	**66.8**	**89.6**	**99.1**	37.1	57.8	80.3	30.2	43.1	45.3
			+ImpartTail (Ours)	**38.8** (+68.0%)	**45.5** (+74.3%)	**45.5** (+73.7%)	60.2 (-9.9%)	82.9 (-7.5%)	98.4 (-0.7%)	**59.4** (+60.1%)	**77.8** (+34.6%)	**92.5** (+15.2%)	**43.6** (+44.4%)	**60.7** (+40.8%)	**63.1** (+39.3%)
		Gaussian Blur	STTran [5]	25.3	28.5	28.6	**67.9**	**89.2**	99.0	38.2	57.6	79.9	30.6	43.1	44.9
			+ImpartTail (Ours)	**38.7** (+53.0%)	**45.7** (+60.4%)	**45.9** (+60.5%)	65.0 (-4.3%)	85.7 (-3.9%)	**99.0** (0.0%)	**59.5** (+55.8%)	**78.1** (+35.6%)	**92.5** (+15.8%)	**43.0** (+40.5%)	**62.0** (+43.9%)	**64.5** (+43.7%)
		Defocus Blur	STTran [5]	25.8	29.2	29.3	**68.4**	**89.6**	**99.1**	39.0	58.1	80.2	31.4	44.1	46.0
			+ImpartTail (Ours)	**38.9** (+50.8%)	**46.0** (+57.5%)	**46.2** (+57.7%)	65.2 (-4.7%)	86.0 (-4.0%)	99.0 (-0.1%)	**60.0** (+53.8%)	**78.5** (+35.1%)	**92.8** (+15.7%)	**43.6** (+38.9%)	**63.2** (+43.3%)	**65.9** (+43.3%)
		Fog	STTran [5]	26.5	30.2	30.3	**70.2**	**91.1**	**99.1**	41.6	61.0	80.5	33.2	46.8	48.7
			+ImpartTail (Ours)	**42.6** (+60.8%)	**50.9** (+68.5%)	**51.1** (+68.6%)	64.8 (-7.7%)	86.3 (-5.3%)	98.8 (-0.3%)	**63.8** (+53.4%)	**80.2** (+31.5%)	**92.7** (+15.2%)	**46.2** (+39.2%)	**65.5** (+40.0%)	**68.2** (+40.0%)
		Frost	STTran [5]	25.6	29.2	29.2	**69.4**	**90.7**	**99.1**	41.0	60.9	80.5	32.7	46.1	48.0
			+ImpartTail (Ours)	**41.0** (+60.2%)	**49.0** (+67.8%)	**49.2** (+68.5%)	62.2 (-10.4%)	84.3 (-7.1%)	98.5 (-0.6%)	**62.5** (+52.4%)	**78.6** (+29.1%)	**92.7** (+15.2%)	**45.1** (+37.9%)	**62.9** (+36.4%)	**65.1** (+35.6%)
		Spatter	STTran [5]	25.7	29.2	29.3	**69.5**	**91.1**	**99.2**	40.0	60.0	80.3	32.0	45.0	47.0
			+ImpartTail (Ours)	**41.3** (+60.7%)	**48.8** (+67.1%)	**48.9** (+66.9%)	58.8 (-15.4%)	82.3 (-9.7%)	98.0 (-1.2%)	**62.0** (+55.0%)	**78.6** (+31.0%)	**92.5** (+15.2%)	**44.6** (+39.4%)	**62.0** (+37.8%)	**64.3** (+36.8%)
		Contrast	STTran [5]	20.5	22.9	23.0	**64.9**	**87.9**	99.0	32.4	53.0	79.6	26.5	36.9	38.5
			+ImpartTail (Ours)	**37.7** (+83.9%)	**44.2** (+93.0%)	**44.3** (+92.6%)	63.3 (-2.5%)	85.0 (-3.3%)	**99.0** (0.0%)	**56.4** (+74.1%)	**76.0** (+43.4%)	**92.0** (+15.6%)	**41.2** (+55.5%)	**57.6** (+56.1%)	**59.6** (+54.8%)
		Brightness	STTran [5]	28.2	32.0	32.1	**71.3**	**91.6**	**99.2**	42.8	62.0	80.4	34.5	49.0	51.2
			+ImpartTail (Ours)	**42.3** (+50.0%)	**50.4** (+57.5%)	**50.5** (+57.3%)	65.9 (-7.6%)	87.2 (-4.8%)	98.9 (-0.3%)	**64.0** (+49.5%)	**80.8** (+30.3%)	**92.8** (+15.4%)	**46.9** (+35.9%)	**67.8** (+38.4%)	**71.0** (+38.7%)
		Pixelate	STTran [5]	24.9	27.9	27.9	**67.3**	**89.4**	**99.1**	37.5	57.5	80.0	30.5	42.9	44.9
			+ImpartTail (Ours)	**38.4** (+54.2%)	**45.5** (+63.1%)	**45.7** (+63.8%)	63.3 (-5.9%)	84.9 (-5.0%)	98.9 (-0.2%)	**59.6** (+58.9%)	**78.3** (+36.2%)	**92.5** (+15.6%)	**43.6** (+43.0%)	**61.9** (+44.3%)	**64.3** (+43.2%)
		Compression	STTran [5]	23.0	25.8	25.8	**66.0**	**88.0**	99.0	35.0	54.6	79.8	27.8	38.9	40.6
			+ImpartTail (Ours)	**36.4** (+58.3%)	**42.2** (+63.6%)	**42.3** (+64.0%)	63.8 (-3.3%)	85.1 (-3.3%)	**99.0** (0.0%)	**54.9** (+56.9%)	**75.8** (+38.8%)	**92.4** (+15.8%)	**41.0** (+47.5%)	**58.6** (+50.6%)	**61.7** (+52.0%)
		Sun Glare	STTran [5]	22.5	25.1	25.2	**66.7**	**89.9**	**99.1**	36.7	56.7	80.0	28.9	40.5	42.3
			+ImpartTail (Ours)	**40.2** (+78.7%)	**47.5** (+89.2%)	**47.7** (+89.3%)	57.9 (-13.2%)	81.7 (-9.1%)	98.0 (-1.1%)	**60.3** (+64.3%)	**77.5** (+36.7%)	**92.7** (+15.9%)	**43.3** (+49.8%)	**59.3** (+46.4%)	**61.0** (+44.2%)

6.4. Robust Scene Graph Anticipation

6.4.1. Findings

Table 30, present the Robustness Evaluation Results for SGCLS for methods STTran+, DSGDetr+, STTran++, DSGDetr++ for Scene Graph Anticipation under various corruption scenarios. The results measure mR@10, mR@20, and mR@50, focusing on the impact of IMPARTAIL across different noise types. For Gaussian Noise, STTran++, mR@10 improves from 5.9 to 9.4 (+59.3%); for DSGDet++, mR@20 improves from 5.7 to 8.8 (+44.4%). However, for STTran+ IMPARTAIL underperforms for all metrics. The same can be observed for Dust, Spatter, Frost and Impulse noises; for all other corruptions, IMPARTAIL outperforms existing methods with the highest increments seen for STTran++ with an average of 40% higher metrics.

6.4.2. Results

Table 30. Robustness Evaluation Results for SGA.

$\mathcal{F}$	Mode	Corruption	Method	With Constraint						No Constraint					
				R@10	R@20	R@50	mR@10	mR@20	mR@50	R@10	R@20	R@50	mR@10	mR@20	mR@50
0.5	sgcls	Gaussian Noise	STTran+ [30]	13.2	13.5	13.5	7.9	8.4	8.4	22.1	27.2	28.7	13.1	18.9	25.7
			+ImparTail (Ours)	7.9 (-40.2%)	8.1 (-40.0%)	8.1 (-40.0%)	5.1 (-35.4%)	5.4 (-35.7%)	5.4 (-35.7%)	12.0 (-45.7%)	18.1 (-33.5%)	22.8 (-20.6%)	7.0 (-46.6%)	11.7 (-38.1%)	19.8 (-23.0%)
			DSGDetr+ [30]	10.1	10.4	10.4	5.5	5.8	5.8	15.2	21.7	24.4	8.2	13.6	22.2
			+ImparTail (Ours)	11.2 (+10.9%)	11.4 (+9.6%)	11.4 (+9.6%)	7.5 (+36.4%)	8.0 (+37.9%)	8.0 (+37.9%)	14.8 (-2.6%)	20.0 (-7.8%)	23.1 (-5.3%)	9.7 (+18.3%)	14.5 (+6.6%)	21.6 (-2.7%)
			STTran++ [30]	10.1	10.5	10.5	5.9	6.4	6.4	15.2	21.1	23.6	8.5	13.9	21.4
			+ImparTail (Ours)	10.0 (-1.0%)	10.3 (-1.9%)	10.3 (-1.9%)	9.4 (+59.3%)	10.2 (+59.4%)	10.2 (+59.4%)	14.5 (-4.6%)	20.1 (-4.7%)	23.3 (-1.3%)	11.4 (+34.1%)	16.3 (+17.3%)	23.0 (+7.5%)
			DSGDetr++ [30]	13.1	13.5	13.5	5.7	6.1	6.1	17.3	22.2	24.7	8.1	14.0	22.6
			+ImparTail (Ours)	12.3 (-6.1%)	12.6 (-6.7%)	12.6 (-6.7%)	8.3 (+45.6%)	8.8 (+44.3%)	8.8 (+44.3%)	15.8 (-8.7%)	21.3 (-4.1%)	24.3 (-1.6%)	10.2 (+25.9%)	15.8 (+12.9%)	22.8 (+0.9%)
		Shot Noise	STTran+ [30]	10.7	11.1	11.1	5.8	6.2	6.2	16.5	22.8	25.6	8.5	14.2	24.7
			+ImparTail (Ours)	8.4 (-21.5%)	8.7 (-21.6%)	8.7 (-21.6%)	5.7 (-1.7%)	6.1 (-1.6%)	6.1 (-1.6%)	12.5 (-24.2%)	18.9 (-17.1%)	23.8 (-7.0%)	7.5 (-11.8%)	12.8 (-9.9%)	20.8 (-15.8%)
			DSGDetr+ [30]	10.6	11.0	11.0	5.9	6.2	6.2	15.8	22.7	25.3	8.7	14.0	22.5
			+ImparTail (Ours)	11.9 (+12.3%)	12.3 (+11.8%)	12.3 (+11.8%)	8.1 (+37.3%)	8.6 (+38.7%)	8.6 (+38.7%)	15.8 (0.0%)	21.4 (-5.7%)	25.2 (-0.4%)	10.4 (+19.5%)	16.0 (+14.3%)	23.5 (+4.4%)
			STTran++ [30]	10.4	10.9	10.9	6.6	7.2	7.2	15.7	21.8	24.1	9.3	14.9	23.1
			+ImparTail (Ours)	10.4 (0.0%)	10.7 (-1.8%)	10.7 (-1.8%)	9.3 (+40.9%)	10.0 (+38.9%)	10.0 (+38.9%)	14.9 (-5.1%)	20.6 (-5.5%)	23.9 (-0.8%)	11.4 (+22.6%)	16.9 (+13.4%)	23.5 (+1.7%)
			DSGDetr++ [30]	13.7	14.1	14.1	6.3	6.7	6.7	17.7	22.8	25.4	8.7	14.4	22.8
			+ImparTail (Ours)	12.7 (-7.3%)	13.1 (-7.1%)	13.1 (-7.1%)	8.4 (+33.3%)	8.8 (+31.3%)	8.8 (+31.3%)	16.4 (-7.3%)	22.2 (-2.6%)	25.3 (-0.4%)	10.3 (+18.4%)	16.3 (+13.2%)	24.1 (+5.7%)
		Impulse Noise	STTran+ [30]	9.7	10.2	10.2	5.1	5.6	5.6	14.8	20.0	22.1	7.6	13.4	21.4
			+ImparTail (Ours)	7.1 (-26.8%)	7.2 (-29.4%)	7.2 (-29.4%)	4.7 (-7.8%)	4.9 (-12.5%)	4.9 (-12.5%)	10.6 (-28.4%)	15.9 (-20.5%)	19.9 (-10.0%)	6.1 (-19.7%)	10.2 (-23.9%)	17.6 (-17.8%)
			DSGDetr+ [30]	10.1	10.5	10.5	5.1	5.6	5.6	14.7	21.0	23.6	7.5	12.7	21.6
			+ImparTail (Ours)	9.4 (-6.9%)	9.7 (-7.6%)	9.7 (-7.6%)	6.1 (+19.6%)	6.6 (+17.9%)	6.6 (+17.9%)	12.9 (-12.2%)	17.3 (-17.6%)	20.3 (-14.0%)	8.1 (+8.0%)	12.1 (-4.7%)	19.1 (-11.6%)
			STTran++ [30]	10.2	10.6	10.6	5.8	6.3	6.3	15.4	20.9	22.8	8.5	14.1	22.0
			+ImparTail (Ours)	10.1 (-1.0%)	10.3 (-2.8%)	10.3 (-2.8%)	8.5 (+46.6%)	8.9 (+41.3%)	8.9 (+41.3%)	14.2 (-7.8%)	19.3 (-7.7%)	22.1 (-3.1%)	10.4 (+22.4%)	15.4 (+9.2%)	21.4 (-2.7%)
			DSGDetr++ [30]	13.3	13.7	13.7	5.7	6.1	6.1	17.0	21.7	23.8	8.2	13.5	20.8
			+ImparTail (Ours)	12.0 (-9.8%)	12.4 (-9.5%)	12.4 (-9.5%)	7.1 (+24.6%)	7.7 (+26.2%)	7.7 (+26.2%)	15.0 (-11.8%)	20.2 (-6.9%)	22.9 (-3.8%)	9.1 (+11.0%)	14.5 (+7.4%)	21.0 (+1.0%)
		Speckle Noise	STTran+ [30]	11.4	11.9	11.9	6.8	7.5	7.5	17.7	24.7	27.6	10.2	17.0	27.8
			+ImparTail (Ours)	9.8 (-14.0%)	10.1 (-15.1%)	10.1 (-15.1%)	7.0 (+2.9%)	7.5 (0.0%)	7.5 (0.0%)	14.6 (-17.5%)	21.9 (-11.3%)	27.8 (+0.7%)	9.2 (-9.8%)	15.7 (-7.6%)	24.9 (-10.4%)
			DSGDetr+ [30]	12.5	13.2	13.2	7.1	7.8	7.8	18.8	26.8	30.2	10.6	17.5	28.1
			+ImparTail (Ours)	13.6 (+8.8%)	13.9 (+5.3%)	13.9 (+5.3%)	9.7 (+36.6%)	10.4 (+33.3%)	10.4 (+33.3%)	18.1 (-3.7%)	24.6 (-8.2%)	28.9 (-4.3%)	12.6 (+18.9%)	19.3 (+10.3%)	28.1 (0.0%)
			STTran++ [30]	12.1	12.7	12.7	7.6	8.3	8.3	18.8	26.0	28.7	11.2	17.9	28.3
			+ImparTail (Ours)	12.1 (0.0%)	12.5 (-1.6%)	12.5 (-1.6%)	10.5 (+38.2%)	11.3 (+36.1%)	11.3 (+36.1%)	17.1 (-9.0%)	23.4 (-10.0%)	27.4 (-4.5%)	12.8 (+14.3%)	19.1 (+6.7%)	27.7 (-2.1%)
			DSGDetr++ [30]	16.1	16.6	16.6	7.5	8.1	8.1	20.9	26.8	29.9	10.5	18.1	28.4
			+ImparTail (Ours)	14.5 (-9.9%)	15.1 (-9.0%)	15.1 (-9.0%)	9.9 (+32.0%)	10.9 (+34.6%)	10.9 (+34.6%)	18.6 (-11.0%)	25.4 (-5.2%)	29.6 (-1.0%)	12.3 (+17.1%)	19.9 (+9.9%)	28.9 (+1.8%)
		Gaussian Blur	STTran+ [30]	16.1	16.9	16.9	9.6	10.6	10.6	24.5	33.9	38.3	13.7	21.9	37.2
			+ImparTail (Ours)	13.7 (-14.9%)	14.2 (-16.0%)	14.2 (-16.0%)	10.5 (+9.4%)	11.1 (+4.7%)	11.1 (+4.7%)	19.9 (-18.8%)	29.7 (-12.4%)	38.0 (-0.8%)	12.9 (-5.8%)	21.0 (-4.1%)	35.9 (-3.5%)
			DSGDetr+ [30]	19.1	20.2	20.2	11.1	12.4	12.4	28.1	39.8	44.8	16.4	26.2	42.0
			+ImparTail (Ours)	19.6 (+2.6%)	20.2 (0.0%)	20.2 (0.0%)	14.4 (+29.7%)	15.6 (+25.8%)	15.6 (+25.8%)	25.6 (-8.9%)	35.3 (-11.3%)	42.6 (-4.9%)	18.1 (+10.4%)	27.1 (+3.4%)	41.7 (-0.7%)
			STTran++ [30]	16.9	17.9	17.9	11.1	12.3	12.3	25.8	35.2	38.6	16.1	24.5	37.8
			+ImparTail (Ours)	17.3 (+2.4%)	17.8 (-0.6%)	17.8 (-0.6%)	15.6 (+40.5%)	16.8 (+36.6%)	16.8 (+36.6%)	23.9 (-7.4%)	32.8 (-6.8%)	37.7 (-2.3%)	18.2 (+13.0%)	26.5 (+8.2%)	38.5 (+1.9%)
			DSGDetr++ [30]	26.0	27.1	27.1	12.4	13.6	13.6	32.6	41.1	45.5	16.3	26.4	42.4
			+ImparTail (Ours)	21.6 (-16.9%)	22.5 (-17.0%)	22.5 (-17.0%)	14.9 (+20.2%)	16.3 (+19.9%)	16.3 (+19.9%)	27.6 (-15.3%)	37.2 (-9.5%)	42.9 (-5.7%)	17.9 (+9.8%)	28.2 (+6.8%)	42.7 (+0.7%)
		Defocus Blur	STTran+ [30]	16.4	17.2	17.2	9.7	10.6	10.6	24.5	34.0	38.4	13.5	21.6	36.4
			+ImparTail (Ours)	13.7 (-16.5%)	14.2 (-17.4%)	14.2 (-17.4%)	10.3 (+6.2%)	10.9 (+2.8%)	10.9 (+2.8%)	20.0 (-18.4%)	29.9 (-12.1%)	38.2 (-0.5%)	12.6 (-6.7%)	21.0 (-2.8%)	36.3 (-0.3%)
			DSGDetr+ [30]	18.8	20.0	20.0	10.8	12.1	12.1	27.6	39.0	44.0	16.0	25.6	41.8
			+ImparTail (Ours)	19.6 (+4.3%)	20.2 (+1.0%)	20.2 (+1.0%)	14.3 (+32.4%)	15.6 (+28.9%)	15.6 (+28.9%)	25.6 (-7.2%)	35.4 (-9.2%)	42.7 (-3.0%)	18.0 (+12.5%)	27.0 (+5.5%)	41.8 (0.0%)
			STTran++ [30]	17.1	18.0	18.0	11.2	12.2	12.2	25.8	35.0	38.4	15.9	24.1	35.8
			+ImparTail (Ours)	17.4 (+1.8%)	17.9 (-0.6%)	17.9 (-0.6%)	15.2 (+35.7%)	16.6 (+36.1%)	16.6 (+36.1%)	23.9 (-7.4%)	32.7 (-6.6%)	37.5 (-2.3%)	17.9 (+12.6%)	26.1 (+8.3%)	38.0 (+6.1%)
			DSGDetr++ [30]	25.1	26.2	26.2	12.0	13.1	13.1	31.6	40.0	44.4	15.9	25.6	41.5
			+ImparTail (Ours)	22.0 (-12.4%)	22.9 (-12.6%)	22.9 (-12.6%)	14.7 (+22.5%)	16.3 (+24.4%)	16.3 (+24.4%)	28.0 (-11.4%)	37.6 (-6.0%)	43.3 (-2.5%)	17.8 (+11.9%)	28.0 (+9.4%)	43.1 (+3.9%)
		Fog	STTran+ [30]	16.9	17.7	17.7	9.8	10.6	10.6	25.9	36.0	40.7	14.2	23.4	38.8
			+ImparTail (Ours)	15.4 (-8.9%)	16.0 (-9.6%)	16.0 (-9.6%)	11.4 (+16.3%)	12.2 (+15.1%)	12.2 (+15.1%)	22.2 (-14.3%)	32.5 (-9.7%)	41.5 (+2.0%)	13.9 (-2.1%)	22.6 (-3.4%)	38.1 (-1.8%)
			DSGDetr+ [30]	19.7	20.8	20.8	10.8	12.0	12.0	29.0	41.0	46.1	16.2	26.4	42.5
			+ImparTail (Ours)	21.6 (+9.6%)	22.3 (+7.2%)	22.3 (+7.2%)	15.2 (+40.7%)	16.6 (+38.3%)	16.6 (+38.3%)	27.7 (-4.5%)	38.1 (-7.1%)	45.6 (-1.1%)	18.8 (+16.0%)	28.7 (+8.7%)	43.2 (+1.6%)
			STTran++ [30]	18.1	19.1	19.1	11.7	12.9	12.9	27.4	37.7	41.7	16.9	26.1	39.3
			+ImparTail (Ours)	19.1 (+5.5%)	19.8 (+3.7%)	19.8 (+3.7%)	15.8 (+35.0%)	17.4 (+34.9%)	17.4 (+34.9%)	26.2 (-4.4%)	35.8 (-5.0%)	41.3 (-1.0%)	18.8 (+11.2%)	28.1 (+7.7%)	40.2 (+2.3%)
			DSGDetr++ [30]	26.5	27.6	27.6	12.7	13.8	13.8	33.2	42.0	46.4	16.6	26.6	42.1
			+ImparTail (Ours)	23.8 (-10.2%)	24.7 (-10.5%)	24.7 (-10.5%)	15.8 (+24.4%)	17.4 (+26.1%)	17.4 (+26.1%)	29.9 (-9.9%)	39.9 (-5.0%)	45.9 (-1.1%)	18.9 (+13.9%)	29.9 (+12.4%)	43.0 (+2.1%)
		Frost	STTran+ [30]	13.6	14.4	14.4	8.2	9.0	9.0	20.3	28.3	32.2	11.6	18.7	33.4
			+ImparTail (Ours)	11.1 (-18.4%)	11.4 (-20.8%)	11.4 (-20.8%)	8.3 (+1.2%)	8.7 (-3.3%)	8.7 (-3.3%)	15.7 (-22.7%)	23.4 (-17.3%)	30.2 (-6.2%)	9.9 (-14.7%)	16.9 (-9.6%)	31.3 (-6.3%)
			DSGDetr+ [30]	15.4	16.3	16.3	9.0	9.9	9.9	22.7	32.0	36.2	13.1	21.3	36.0
			+ImparTail (Ours)	16.3 (+5.8%)	16.7 (+2.5%)	16.7 (+2.5%)	12.4 (+37.8%)	13.3 (+34.3%)	13.3 (+34.3%)	21.2 (-6.6%)	29.3 (-8.4%)	35.0 (-3.3%)	15.0 (+14.5%)	23.2 (+8.9%)	34.7 (-3.6%)
			STTran++ [30]	14.1	14.8	14.8	9.6	10.5	10.5	21.1	29.0	32.1	13.3	20.9	32.4
			+ImparTail (Ours)	13.8 (-2.1%)	14.2 (-4.1%)	14.2 (-4.1%)	13.8 (+43.7%)	15.3 (+45.7%)	15.3 (+45.7%)	18.6 (-11.8%)	26.1 (-10.0%)	30.4 (-5.3%)	15.8 (+18.8%)	23.4 (+12.0%)	32.5 (+0.3%)
			DSGDetr++ [30]	20.5	21.3	21.3	9.9	10.8	10.8	25.8	32.8	36.7	12.9	21.3	34.4
			+ImparTail (Ours)	18.4 (-10.2%)	19.1 (-10.3%)	19.1 (-10.3%)	13.2 (+33.3%)	14.4 (+33.3%)	14.4 (+33.3%)	23.1 (-10.5%)	31.5 (-4.0%)	36.4 (-0.8%)	15.4 (+19.4%)	24.6 (+15.5%)	37.7 (+9.6%)
		Spatter	STTran+ [30]	15.6	16.5	16.5	8.5	9.4	9.4	23.6	32.7	36.8	12.3	20.1	35.5
			+ImparTail (Ours)	12.9 (-17.3%)	13.4 (-18.8%)	13.4 (-18.8%)	8.5 (0.0%)	8.9 (-5.3%)	8.9 (-5.3%)	18.6 (-21.2%)	28.0 (-14.4%)	36.0 (-2.2%)	10.8 (-12.2%)	18.2 (-9.5%)	33.4 (-5.9%)
			DSGDetr+ [30]	18.2	19.3	19.3	9.8	10.9	10.9	26.8	37.8	42.5	14.3	23.6	39.6
			+ImparTail (Ours)	20.0 (+9.9%)	20.7 (+7.3%)	20.7 (+7.3%)	12.4 (+26.5%)	13.8 (+26.6%)	13.8 (+26.6%)	25.1 (-6.3%)	34.5 (-8.7%)	41.1 (-3.3%)	15.7 (+9.8%)	24.5 (+3.8%)	38.0 (-4.0%)
			STTran++ [30]	16.4	17.3	17.3	10.0	11.1	11.1	24.6	33.6	36.9	14.6	22.3	34.6
			+ImparTail (Ours)	16.4 (0.0%)	17.0 (-1.7%)	17.0 (-1.7%)	13.9 (+39.0%)	15.2 (+36.9%)	15.2 (+36.9%)	22.6 (-8.1%)	31.1 (-7.4%)	36.4 (-1.4%)	16.4 (+12.3%)	24.0 (+7.6%)	35.6 (+2.9%)
			DSGDetr++ [30]	23.9	24.9	24.9	10.9	11.9	11.9	29.6	37.4	41.4	14.3	23.5	37.8
			+ImparTail (Ours)	22.5 (-5.9%)	23.4 (-6.0%)	23.4 (-6.0%)	14.1 (+29.4%)	15.4 (+29.4%)	15.4 (+29.4%)	27.9 (-5.7%)	37.4 (0.0%)	43.0 (+3.9%)	16.8 (+17.5%)	26.8 (+14.0%)	39.8 (+5.3%)
		Contrast	STTran+ [30]	14.9	15.6	15.6	8.4	9.1	9.1	22.7	31.5	35.4	12.0	20.0	33.0
			+ImparTail (Ours)	12.1 (-18.8%)	12.6 (-19.2%)	12.6 (-19.2%)	9.2 (+9.5%)	9.8 (+7.7%)	9.8 (+7.7%)	17.9 (-21.1%)	27.2 (-13.7%)	35.2 (-0.6%)	11.4 (-5.0%)	19.3 (-3.5%)	34.1 (+3.3%)
			DSGDetr+ [30]	17.3	18.2	18.2	9.4	10.2	10.2	25.5	36.4	40.7	13.9	23.1	36.0
			+ImparTail (Ours)	19.3 (+11.6%)	19.8 (+8.8%)	19.8 (+8.8%)	13.6 (+44.7%)	14.7 (+44.1%)	14.7 (+44.1%)	24.9 (-2.4%)	34.6 (-4.9%)	41.1 (+1.0%)	16.5 (+18.7%)	25.7 (+11.3%)	38.6 (+7.2%)
			STTran++ [30]	15.9	16.8	16.8	10.2	11.3	11.3	24.0	33.2	36.8	14.4	22.9	36.1
			+ImparTail (Ours)	16.0 (+0.6%)	16.5 (-1.8%)	16.5 (-1.8%)	13.5 (+32.4%)	14.8 (+31.0%)	14.8 (+31.0%)	21.8 (-9.2%)	30.3 (-8.7%)	35.2 (-4.3%)	16.0 (+11.1%)	23.8 (+3.9%)	33.4 (-7.5%)
			DSGDetr++ [30]	23.6	24.6	24.6	11.0	11.9	11.9	29.8	38.0	42.1	14.5	24.0	37.5
			+ImparTail (Ours)	20.7 (-12.3%)	21.5 (-12.6%)	21.5 (-12.6%)	14.4 (+30.9%)	15.6 (+31.1%)	15.6 (+31.1%)	26.3 (-11.7%)	35.7 (-6.1%)	41.2 (-2.1%)	16.8 (+15.9%)	26.5 (+10.4%)	38.5 (+2.7%)
		Brightness	STTran+ [30]	19.6	20.7	20.7	11.0	12.1	12.1	29.9	41.2	46.4	15.9	26.1	44.1
			+ImparTail (Ours)	16.9 (-13.8%)	17.5 (-15.5%)	17.5 (-15.5%)	11.7 (+6.4%)	12.5 (+3.3%)	12.5 (+3.3%)	24.1 (-19.4%)	35.7 (-13.3%)	45.7 (-1.5%)	14.9 (-6.3%)	24.5 (-6.1%)	42.0 (-4.8%)
			DSGDetr+ [30]	21.9	23.2	23.2	12.2	13.5	13.5	32.1	45.1	50.5	18.0	29.1	47.1
			+ImparTail (Ours)	23.2 (+5.9%)	23.9 (+3.0%)	23.9 (+3.0%)	15.9 (+30.3%)	17.2 (+27.4%)	17.2 (+27.4%)	29.7 (-7.5%)	40.9 (-9.3%)	48.7 (-3.6%)	19.8 (+10.0%)	29.8 (+2.4%)	46.0 (-2.3%)
			STTran++ [30]	20.2	21.3	21.3	12.8	14.1	14.1	30.3	41.8	46.0	18.3	28.7	43.9
			+ImparTail (Ours)	21.0 (+4.0%)	21.8 (+2.3%)	21.8 (+2.3%)	17.7 (+38.3%)	19.3 (+36.9%)	19.3 (+36.9%)	28.7 (-5.3%)	39.6 (-5.3%)	45.6 (-0.9%)	20.8 (+13.7%)	30.6 (+6.6%)	43.4 (-1.1%)
			DSGDetr++ [30]	28.9	30.0	30.0	13.6	14.7	14.7	36.0	45.2	49.6	17.8	29.0	46.3
			+ImparTail (Ours)	25.7 (-11.1%)	26.7 (-11.0%)	26.7 (-11.0%)	16.6 (+22.1%)	18.1 (+23.1%)	18.1 (+23.1%)	32.4 (-10.0%)	43.5 (-3.8%)	49.7 (+0.2%)	20.2 (+13.5%)	32.0 (+10.3%)	47.9 (+3.5%)
		Pixelate	STTran+ [30]	18.5	19.5	19.5	10.4	11.4	11.4	28.0	38.8	43.8	14.8	24.3	41.5
			+ImparTail (Ours)	15.7 (-15.1%)	16.2 (-16.9%)	16.2 (-16.9%)	11.9 (+14.4%)	12.7 (+11.4%)	12.7 (+11.4%)	22.8 (-18.6%)	34.4 (-11.3%)	44.1 (+0.7%)	14.8 (0.0%)	24.4 (+0.4%)	40.8 (-1.7%)
			DSGDetr+ [30]	20.9	22.2	22.2	11.6	12.8	12.8	30.9	43.6	49.0	17.2	27.4	45.8
			+ImparTail (Ours)	21.9 (+4.8%)	22.7 (+2.3%)	22.7 (+2.3%)	15.7 (+35.3%)	16.9 (+32.0%)	16.9 (+32.0%)	28.5 (-7.8%)	39.8 (-8.7%)	47.7 (-2.7%)	19.1 (+11.0%)	29.4 (+7.3%)	44.2 (-3.5%)
			STTran++ [30]	19.3	20.4	20.4	11.9	13.1	13.1	29.2	40.2	44.3	17.2	27.0	41.6
			+ImparTail (Ours)	19.4 (+0.5%)	20.1 (-1.5%)	20.1 (-1.5%)	15.8 (+32.8%)	17.5 (+33.6%)	17.5 (+33.6%)	26.2 (-10.3%)	36.9 (-8.2%)	42.8 (-3.4%)	18.5 (+7.6%)	27.9 (+3.3%)	40.8 (-1.9%)
			DSGDetr++ [30]	27.2	28.3	28.3	12.6	13.7	13.7	34.0	43.1	47.7	16.2	26.4	43.0
			+ImparTail (Ours)	24.0 (-11.8%)	24.9 (-12.0%)	24.9 (-12.0%)	15.6 (+23.8%)	17.1 (+24.8%)	17.1 (+24.8%)	30.7 (-9.7%)	41.4 (-3.9%)	47.3 (-0.8%)	18.8 (+16.0%)	29.9 (+13.3%)	44.6 (+3.7%)
		Compression	STTran+ [30]	17.3	18.3	18.3	9.3	10.2	10.2	26.1	36.0	40.6	13.1	21.9	37.5
			+ImparTail (Ours)	14.3 (-17.3%)	14.8 (-19.1%)	14.8 (-19.1%)	10.1 (+8.6%)	10.8 (+5.9%)	10.8 (+5.9%)	21.0 (-19.5%)	32.1 (-10.8%)	41.6 (+2.5%)	12.5 (-4.6%)	21.6 (-1.4%)	36.7 (-2.1%)
			DSGDetr+ [30]	19.9	21.0	21.0	10.6	11.8	11.8	29.0	41.4	46.6	15.6	25.5	43.6
			+ImparTail (Ours)	20.7 (+4.0%)	21.4 (+1.9%)	21.4 (+1.9%)	14.5 (+36.8%)	15.9 (+34.7%)	15.9 (+34.7%)	26.7 (-7.9%)	37.6 (-9.2%)	45.3 (-2.8%)	17.8 (+14.1%)	27.8 (+9.0%)	41.6 (-4.6%)
			STTran++ [30]	18.1	19.1	19.1	10.8	12.0	12.0	27.2	37.4	41.2	15.8	24.4	38.5
			+ImparTail (Ours)	18.1 (0.0%)	18.7 (-2.1%)	18.7 (-2.1%)	14.5 (+34.3%)	16.0 (+33.3%)	16.0 (+33.3%)	24.6 (-9.6%)	34.2 (-8.6%)	39.7 (-3.6%)	17.4 (+10.1%)	25.7 (+5.3%)	37.1 (-3.6%)
			DSGDetr++ [30]	26.0	27.1	27.1	11.5	12.5	12.5	32.6	41.2	45.9	14.9	24.5	40.2
			+ImparTail (Ours)	22.9 (-11.9%)	23.9 (-11.8%)	23.9 (-11.8%)	14.5 (+26.1%)	16.0 (+28.0%)	16.0 (+28.0%)	29.2 (-10.4%)	39.3 (-4.6%)	45.2 (-1.5%)	17.4 (+16.8%)	28.3 (+15.5%)	43.0 (+7.0%)
		Sun Glare	STTran+ [30]	12.2	12.8	12.8	7.1	7.7	7.7	18.6	26.0	29.5	10.0	16.3	27.1
			+ImparTail (Ours)	10.2 (-16.4%)	10.5 (-18.0%)	10.5 (-18.0%)	7.5 (+5.6%)	8.0 (+3.9%)	8.0 (+3.9%)	14.5 (-22.0%)	22.2 (-14.6%)	28.8 (-2.4%)	9.2 (-8.0%)	15.9 (-2.5%)	26.9 (-0.7%)
			DSGDetr+ [30]	13.8	14.5	14.5	7.6	8.4	8.4	20.0	29.2	33.1	11.0	18.3	31.0
			+ImparTail (Ours)	14.7 (+6.5%)	15.1 (+4.1%)	15.1 (+4.1%)	10.2 (+34.2%)	11.0 (+31.0%)	11.0 (+31.0%)	18.8 (-6.0%)	25.9 (-11.3%)	30.8 (-6.9%)	12.8 (+16.4%)	19.4 (+6.0%)	30.1 (-2.9%)
			STTran++ [30]	12.3	13.1	13.1	7.7	8.6	8.6	18.7	26.0	29.0	11.0	17.6	27.2
			+ImparTail (Ours)	12.7 (+3.3%)	13.2 (+0.8%)	13.2 (+0.8%)	11.4 (+48.1%)	12.4 (+44.2%)	12.4 (+44.2%)	17.2 (-8.0%)	24.0 (-7.7%)	27.9 (-3.8%)	13.2 (+20.0%)	20.4 (+15.9%)	28.7 (+5.5%)
			DSGDetr++ [30]	18.0	18.6	18.6	8.3	8.9	8.9	22.6	28.7	31.9	10.8	18.1	29.1
			+ImparTail (Ours)	15.8 (-12.2%)	16.4 (-11.8%)	16.4 (-11.8%)	10.8 (+30.1%)	11.7 (+31.5%)	11.7 (+31.5%)	19.9 (-11.9%)	26.8 (-6.6%)	30.7 (-3.8%)	12.7 (+17.6%)	20.1 (+11.0%)	29.9 (+2.7%)
		Dust	STTran+ [30]	12.1	12.7	12.7	7.4	8.0	8.0	18.6	25.7	28.7	10.6	16.8	28.4
			+ImparTail (Ours)	9.2 (-24.0%)	9.5 (-25.2%)	9.5 (-25.2%)	7.2 (-2.7%)	7.7 (-3.8%)	7.7 (-3.8%)	13.8 (-25.8%)	21.2 (-17.5%)	27.6 (-3.8%)	9.0 (-15.1%)	15.4 (-8.3%)	26.4 (-7.0%)
			DSGDetr+ [30]	13.2	14.0	14.0	7.4	8.1	8.1	19.4	27.8	31.6	10.8	17.8	29.1
			+ImparTail (Ours)	13.8 (+4.5%)	14.2 (+1.4%)	14.2 (+1.4%)	10.4 (+40.5%)	11.1 (+37.0%)	11.1 (+37.0%)	18.3 (-5.7%)	25.5 (-8.3%)	30.3 (-4.1%)	12.6 (+16.7%)	19.5 (+9.6%)	30.2 (+3.8%)
			STTran++ [30]	12.7	13.4	13.4	8.4	9.3	9.3	19.0	26.2	29.1	11.9	18.4	28.4
			+ImparTail (Ours)	12.4 (-2.4%)	12.8 (-4.5%)	12.8 (-4.5%)	10.6 (+26.2%)	11.7 (+25.8%)	11.7 (+25.8%)	17.1 (-10.0%)	23.8 (-9.2%)	27.5 (-5.5%)	12.4 (+4.2%)	18.7 (+1.6%)	25.8 (-9.2%)
			DSGDetr++ [30]	16.4	17.1	17.1	8.0	8.7	8.7	20.7	26.8	30.1	10.2	17.4	27.9
			+ImparTail (Ours)	15.7 (-4.3%)	16.2 (-5.3%)	16.2 (-5.3%)	11.0 (+37.5%)	12.1 (+39.1%)	12.1 (+39.1%)	19.7 (-4.8%)	26.7 (-0.4%)	30.9 (+2.7%)	13.0 (+27.5%)	20.5 (+17.8%)	29.7 (+6.5%)
		Saturate	STTran+ [30]	21.3	22.5	22.5	12.0	13.2	13.2	32.4	44.5	49.9	17.5	28.5	47.5
			+ImparTail (Ours)	19.0 (-10.8%)	19.7 (-12.4%)	19.7 (-12.4%)	13.6 (+13.3%)	14.4 (+9.1%)	14.4 (+9.1%)	26.9 (-17.0%)	38.9 (-12.6%)	49.1 (-1.6%)	17.3 (-1.1%)	28.0 (-1.8%)	46.3 (-2.5%)
			DSGDetr+ [30]	23.1	24.4	24.4	13.0	14.4	14.4	34.0	47.6	53.4	19.6	31.3	49.3
			+ImparTail (Ours)	24.6 (+6.5%)	25.3 (+3.7%)	25.3 (+3.7%)	17.0 (+30.8%)	18.2 (+26.4%)	18.2 (+26.4%)	31.7 (-6.8%)	43.5 (-8.6%)	51.9 (-2.8%)	21.8 (+11.2%)	32.9 (+5.1%)	49.4 (+0.2%)
			STTran++ [30]	21.9	23.1	23.1	13.9	15.4	15.4	33.2	45.5	50.0	20.2	31.7	48.1
			+ImparTail (Ours)	23.2 (+5.9%)	24.1 (+4.3%)	24.1 (+4.3%)	19.1 (+37.4%)	21.1 (+37.0%)	21.1 (+37.0%)	31.4 (-5.4%)	42.8 (-5.9%)	49.1 (-1.8%)	23.2 (+14.9%)	33.7 (+6.3%)	48.6 (+1.0%)
			DSGDetr++ [30]	30.8	32.1	32.1	14.6	15.9	15.9	38.5	48.4	53.1	19.2	31.4	49.0
			+ImparTail (Ours)	27.8 (-9.7%)	28.9 (-10.0%)	28.9 (-10.0%)	18.0 (+23.3%)	19.4 (+22.0%)	19.4 (+22.0%)	34.9 (-9.4%)	46.7 (-3.5%)	53.2 (+0.2%)	21.8 (+13.5%)	34.4 (+9.6%)	51.3 (+4.7%)

References

[1] Emanuel Ben Baruch, T. Ridnik, Nadav Zamir, Asaf Noy, Itamar Friedman, Matan Protter, and Lihi Zelnik-Manor. Asymmetric loss for multi-label classification. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 82–91, 2020. 3

[2] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Aréchiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Neural Information Processing Systems*, 2019. 3

[3] N. Chawla, K. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *ArXiv*, abs/1106.1813, 2002. 3

[4] Zuyao Chen, Jinlin Wu, Zhen Lei, Zhaoxiang Zhang, and Changwen Chen. Expanding scene graph boundaries: Fully open-vocabulary scene graph generation via visual-concept alignment and retention, 2023. 2

[5] Yuren Cong, Wentong Liao, H. Ackermann, M. Yang, and B. Rosenhahn. Spatial-temporal transformer for dynamic scene graph generation. *IEEE International Conference on Computer Vision*, 2021. 2, 17, 18, 33

[6] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples, 2019. 3

[7] Chris Drummond and Robert C. Holte. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. 2003. 3

[8] Jeffrey L. Elman. Learning and development in neural networks: the importance of starting small. *Cognition*, 48:71–99, 1993. 3

[9] Shengyu Feng, Hesham Mostafa, Marcel Nassar, Somdeb Majumdar, and Subarna Tripathi. Exploiting long-term dependencies for generating dynamic scene graphs. *IEEE Workshop/Winter Conference on Applications of Computer Vision*, 2021. 17, 18, 32

[10] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, 1997. 3

[11] Alex Graves, Marc G. Bellemare, Jacob Menick, Rémi Munos, and Koray Kavukcuoglu. Automated curriculum learning for neural networks. *ArXiv*, abs/1704.03003, 2017. 3

[12] Guy Hacohen and Daphna Weinshall. On the power of curriculum learning in training deep networks. In *International Conference on Machine Learning*, 2019. 3

[13] Yizeng Han, Yifan Pu, Zihang Lai, Chaofei Wang, Shiji Song, Junfen Cao, Wenhui Huang, Chao Deng, and Gao Huang. Learning to weight samples for dynamic early-exiting networks. *ArXiv*, abs/2209.08310, 2022. 3

[14] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5375–5384, 2016. 3

[15] Radu Tudor Ionescu, Bogdan Alexe, Marius Leordeanu, Marius Claudiu Popescu, Dim P. Papadopoulos, and Vittorio Ferrari. How hard can it be? estimating the difficulty of visual search in an image. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2157–2166, 2016. 3

[16] Jingwei Ji, Ranjay Krishna, Li Fei-Fei, Fei-Fei Li, and Juan Carlos Niebles. Action genome: Actions as composition of spatio-temporal scene graphs. *arXiv: Computer Vision and Pattern Recognition*, 2019. 2

[17] Lu Jiang, Deyu Meng, Teruko Mitamura, and Alexander Hauptmann. Easy samples first: Self-paced reranking for zero-example multimedia search. *Proceedings of the 22nd ACM international conference on Multimedia*, 2014. 3

[18] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International Conference on Machine Learning*, 2017. 3

[19] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. *ArXiv*, abs/1910.09217, 2019. 3

[20] Anant Khandelwal. Correlation debiasing for unbiased scene graph generation in videos, 2023. 2

[21] Kibum Kim, Kanghoon Yoon, Jaehyeong Jeon, Yeonjun In, Jinyoung Moon, Donghyun Kim, and Chanyoung Park. Llm4sgg: Large language model for weakly supervised scene graph generation, 2023. 2

[22] Ue-Hwan Kim, Jin-Man Park, Taek jin Song, and Jong-Hwan Kim. 3-d scene graph: A sparse and semantic representation of physical environments for intelligent agents. *IEEE Transactions on Cybernetics*, 50:4921–4933, 2019. 2

[23] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *Int. J. Comput. Vis.*, 123(1):32–73, 2017. 2

[24] M. Pawan Kumar, Ben Packer, and Daphne Koller. Self-paced learning for latent variable models. In *Neural Information Processing Systems*, 2010. 3

[25] Lin Li, Jun Xiao, Guikun Chen, Jian Shao, Yueting Zhuang, and Long Chen. Zero-shot visual relation detection via composite visual cues from large language models, 2023. 2

[26] Mengke Li, Yiu-Ming Cheung, and Yang Lu. Long-tailed visual recognition via gaussian clouded logit adjustment. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 6919–6928. IEEE, 2022. 3

[27] Tambet Matiisen, Avital Oliver, Taco Cohen, and John Schulman. Teacher–student curriculum learning. *IEEE Transactions on Neural Networks and Learning Systems*, 31:3732–3740, 2017. 3

[28] Sayak Nag, Kyle Min, Subarna Tripathi, and Amit K Roy-Chowdhury. Unbiased scene graph generation in videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22803–22813, 2023. 2, 7, 9

[29] Wanli Ouyang, Xiaogang Wang, Cong Zhang, and Xiaokang Yang. Factors in finetuning deep model for object detection with long-tail distribution. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 864–873, 2016. 3

[30] Rohith Peddi, Saksham Singh, Saurabh, Parag Singla, and Vibhav Gogate. Towards scene graph anticipation. In *Computer Vision – ECCV 2024*, pages 159–175, Cham, 2025. Springer Nature Switzerland. 2, 12, 14, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 34

[31] Mengye Ren, Wenyuan Zeng, Binh Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *International Conference on Machine Learning*, 2018. 3

[32] Andrew I. Schein and Lyle H. Ungar. Active learning for logistic regression. *Machine Learning*, 2005. 3

[33] Xindi Shang, Tongwei Ren, Jingfan Guo, Hanwang Zhang, and Tat-Seng Chua. Video visual relation detection. In *Proceedings of the 25th ACM International Conference on Multimedia*, page 1300–1308, New York, NY, USA, 2017. Association for Computing Machinery. 2

[34] Suprosanna Shit, Rajat Koner, Bastian Wittmann, Johannes Paetzold, Ivan Ezhov, Hongwei Li, Jiazhen Pan, Sahand Sharifzadeh, Georgios Kaissis, Volker Tresp, and Bjoern Menze. Relationformer: A unified framework for image-to-graph generation, 2022. 2

[35] Abhinav Shrivastava, Abhinav Kumar Gupta, and Ross B. Girshick. Training region-based object detectors with online hard example mining. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 761–769, 2016. 3

[36] Jingru Tan, Changbao Wang, Buyu Li, Quanquan Li, Wanli Ouyang, Changqing Yin, and Junjie Yan. Equalization loss for long-tailed object recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11659–11668, 2020. 3

[37] Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella X. Yu. Long-tailed recognition by routing diverse distribution-aware experts. *ArXiv*, abs/2010.01809, 2020. 3

[38] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to model the tail. In *Neural Information Processing Systems*, 2017. 3

[39] Daphna Weinshall and Gad Cohen. Curriculum learning by transfer learning: Theory and experiments with deep networks. In *International Conference on Machine Learning*, 2018. 3

[40] Tong Wu, Qingqiu Huang, Ziwei Liu, Yu Wang, and Dahua Lin. Distribution-balanced loss for multi-label classification in long-tailed datasets. *ArXiv*, abs/2007.09654, 2020. 3

[41] Liuyu Xiang and Guiguang Ding. Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification. In *European Conference on Computer Vision*, 2020. 3

[42] Chong You, Chi Li, Daniel P. Robinson, and René Vidal. A scalable exemplar-based subspace clustering algorithm for class-imbalanced data. In *European Conference on Computer Vision*, 2018. 3

[43] Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5419–5428, 2016. 3

[44] X. Zhang, Z. Wu, Z. Weng, H. Fu, J. Chen, Y. Jiang, and L. S. Davis. Videolt: large-scale long-tailed video recognition. 2021. 3

[45] Shu Zhao and Huijuan Xu. Less is more: Toward zero-shot local scene graph generation via foundation models, 2023. 2

[46] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9716–9725, 2019. 3

[47] Zijian Zhou, Miaojing Shi, and Holger Caesar. Vlprompt: Vision-language prompting for panoptic scene graph generation, 2023. 2