1. Supplementary Material

1.1. Additional Model Pipelines

We illustrate how *Hallucinator* plugged into other models we used in this paper. To be specific, Figure 1 demonstrates the overall pipeline of MoCoV1. The corresponding InfoNCE loss[4] could be defined as:

$$\mathcal{L}_{MoCo} = -\log \frac{\exp(q \cdot k/\tau) + \exp(\hat{q} \cdot k/\tau)}{\sum_{k^-} \exp(q \cdot k^-/\tau) + \exp(q \cdot k/\tau) + \exp(\hat{q} \cdot k/\tau)}, \tag{1}$$

where τ is a temperature hyper-parameter. (q, k) and $(\hat{q}, k)$ are two positive pairs. (q, k^-) are negative pairs. All k^- vectors are stored in a queue structure.

Following this, we demonstrate the pipeline of SimCLR[1] with the proposed *Hallucinator* in Figure 2. Again, (q, k) and $(\hat{q}, k)$ are defined as two positive pairs for consistency. For a positive pair (q, k), we define the loss the same as before[1]:

$$\mathcal{L}_{SimCLR}(q, k) = -\log \frac{\exp(sim(q, k)/\tau)}{\sum_{k^-}^{4N} \mathbb{1}_{[k^- \neq q]} \exp(sim(q, k^-)/\tau)} \tag{2}$$

where $sim(q, k) = q^T k / \|q\| \|k\|$, i.e., cosine similarity. N represents the batch size and τ is a temperature hyper-parameter. Originally, we had $2N$ data points. The number increase to $4N$ as the *Hallucinator* added. $\mathbb{1}_{[k^- \neq q]} \in \{0, 1\}$ is an indicator function, which equals to 1 iff $k^- \neq q$. Notably, we compute the loss for all the positives, including (q, k), (k, q), $(k, \hat{q})$ and $(\hat{q}, k)$.

We continue to demonstrate the structure of SimSiam with *Hallucinator* in Figure 3. Notably, asymmetric extrapolation is applied in embedding space. Therefore, the projector is added before feeding the feature vector k to *Hallucinator*. We define one side of the loss as follows:

$$D(p, k) = -\left(\frac{q}{\|q\|_2} \cdot \frac{k}{\|k\|_2} + \frac{\hat{q}}{\|\hat{q}\|_2} \cdot \frac{k}{\|k\|_2} \right) \tag{3}$$

where (p, k) are encoded from two views with and without a projector h respectively. If the encoder is noted as f, then $p = h(f(view1))$ and $k = f(view2)$. Notably, the total loss is symmetric. If we switch two views by defining $k' = f(view1)$ and $p' = h(f(view2))$. We get the final form of loss as:

$$\mathcal{L}_{SimSiam} = \frac{1}{2} D(p, k) + \frac{1}{2}(p', k') \tag{4}$$

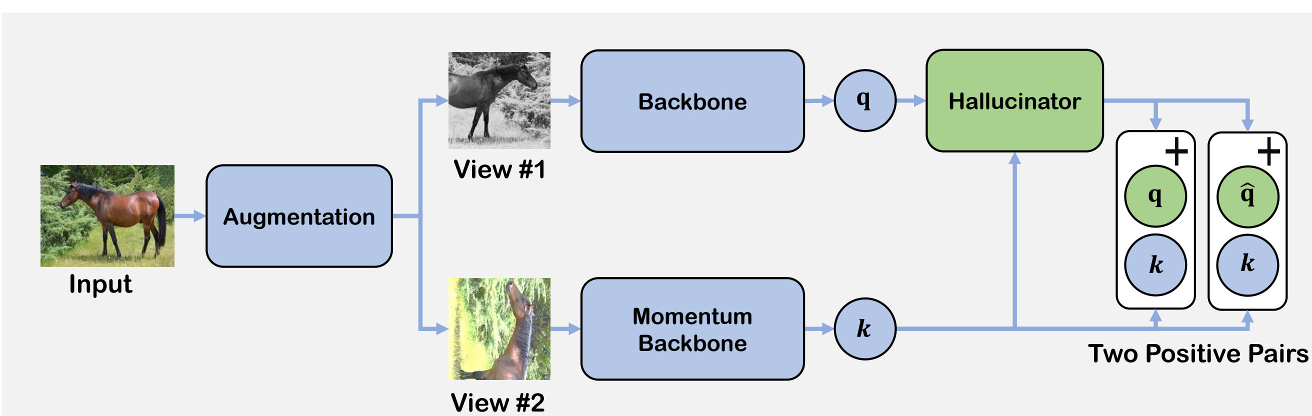

Figure 1. Illustration of the pipeline based on MoCoV1[3] with *Hallucinator*.

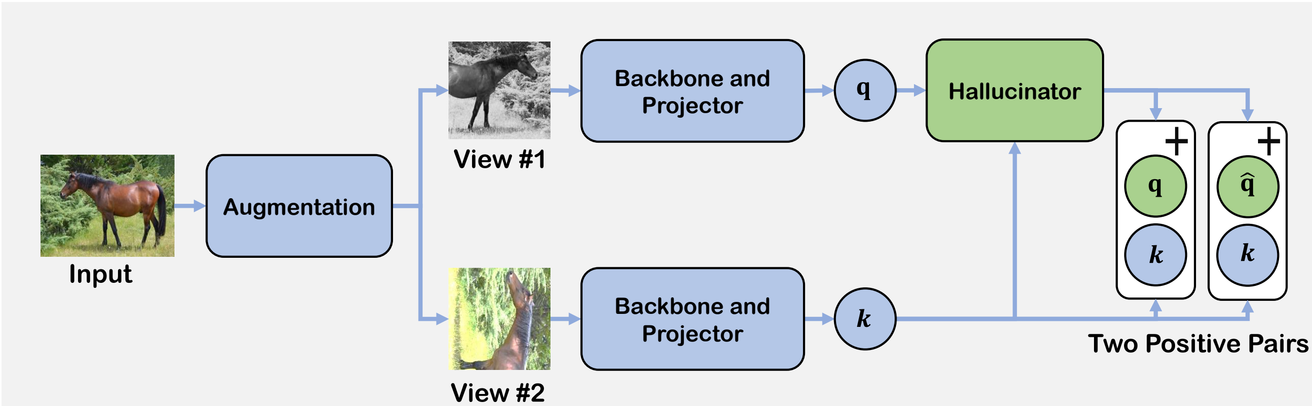

Figure 2. Illustration of the pipeline based on SimCLR[1] with *Hallucinator*.

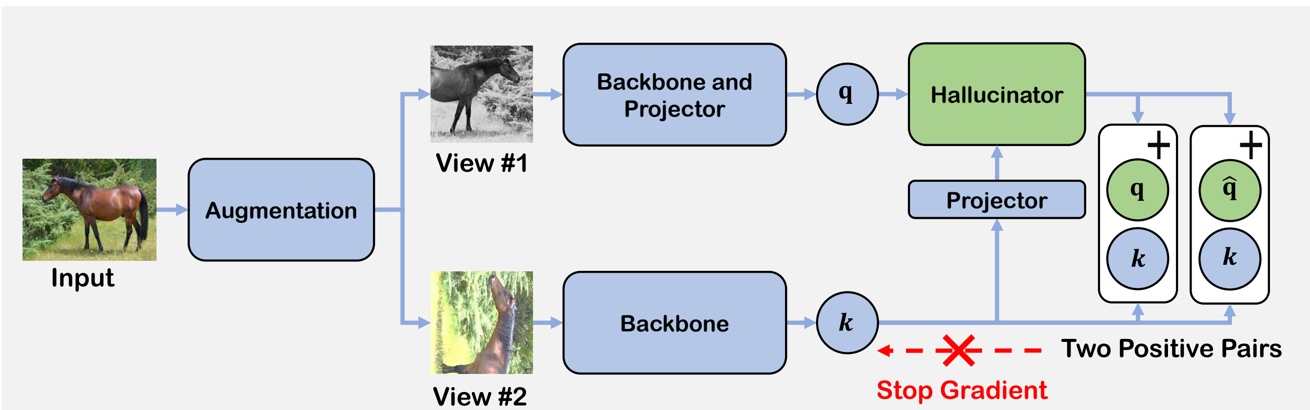

Figure 3. Illustration of the pipeline based on SimSiam[2] with *Hallucinator*.

1.2. The Center Sampling Method

We visualize the different cropping methods in Figure 4. Importantly, random cropping sometimes introduces false positive pairs[5], posing a negative influence on overall training. Such adverse influence is more noticeable when hallucinated sample is generated without sharing mutual semantic meaning with the other positive feature vector. While center cropping reduces the sampling areas and successfully avoids positive pairs, it contains less variance, thus gaining sub-optimal embeddings. If center-suppressed sampling is introduced, the mutual information between positive pairs will be reduced. Then, the benefits of hallucinated samples can be further enhanced.

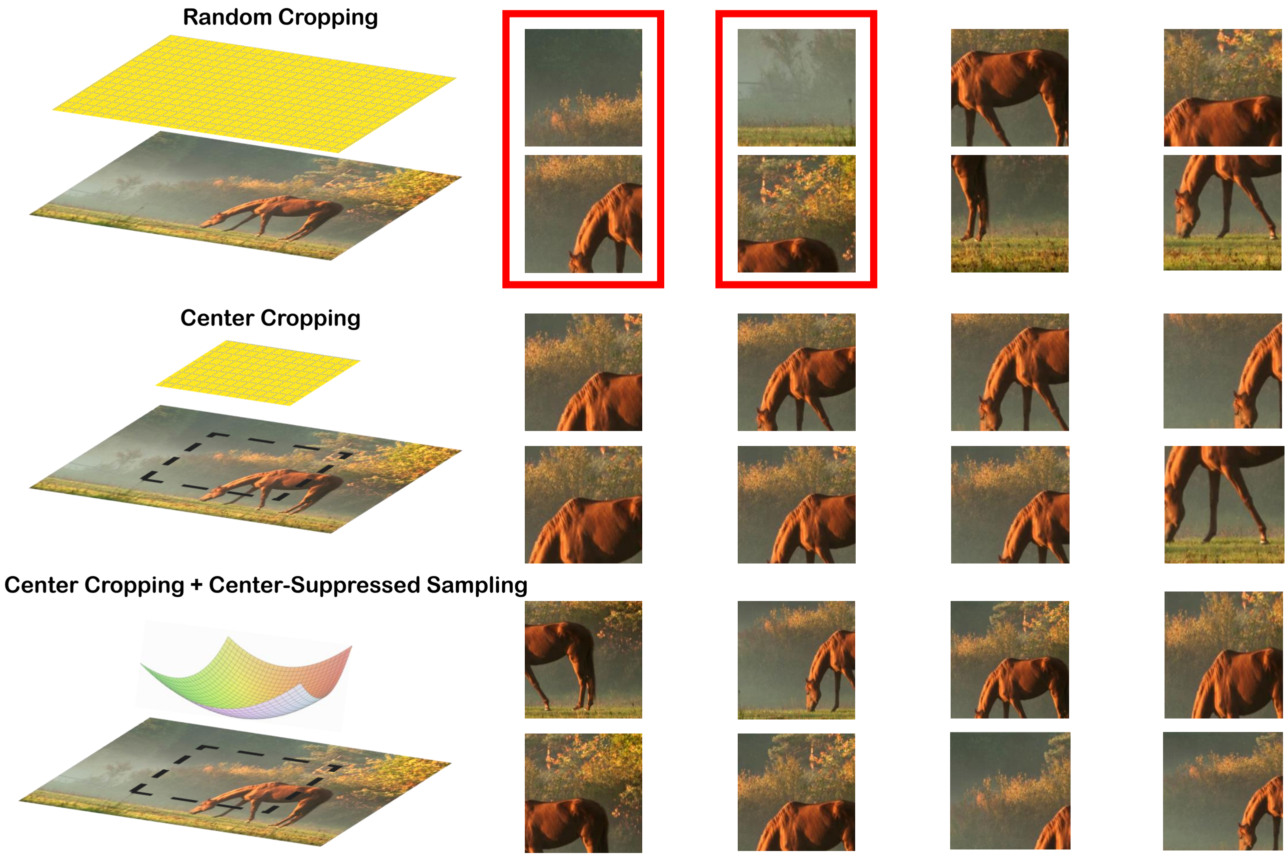

Figure 4. Visualization of three different cropping methods. We demonstrate the sampling probability(distribution) and operable regions for three settings on the left, and their sampled pairs on the right. Pairs cropped by Random Cropping sometimes introduce false positives, which are highlighted in red boxes. Center Cropping avoids false positives but fails to incorporate sufficient variance. Center Cropping with center-suppressed sampling provides the idea views for *Hallucinator* with enough variance and mutual semantic information.

References

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 1, 2

[2] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758, 2021. 2

[3] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 1

[4] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 1

[5] Xiangyu Peng, Kai Wang, Zheng Zhu, Mang Wang, and Yang You. Crafting better contrastive views for siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16031–16040, 2022. 2