Sequential Graph Convolutional Network for Active Learning (Supplementary Material)

Razvan Caramalau[1], Binod Bhattarai[1] and Tae-Kyun Kim[1,2]
[1]Imperial College London, UK
[2]KAIST, South Korea
{ r.caramalau18, b.bhattarai, tk.kim}@imperial.ac.uk

A. Datasets

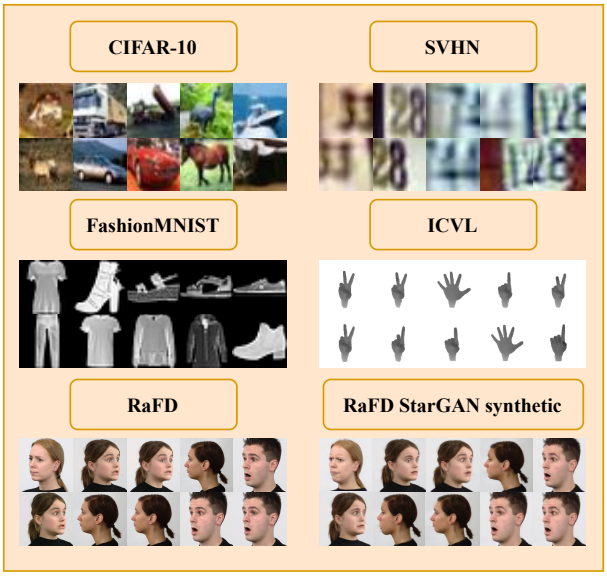

Figure A.1: This Figure shows some of the randomly sampled images from the data sets we use to validate our methods. Effectiveness of our method on these diverse characterstics of datasets demonstrate its generic nature.

Here, we present a extended description of the datasets we used to evaluate our algorithms and the compared baselines. We evaluated our methods together with the others on four challenging image classification benchmarks: CIFAR-10[3], CIFAR-100[3], FashionMNIST[7] and SVHN[2]. Each of the datasets has different properties and present new challenges for the active learning framework. FashionMNIST is a grey scale image dataset. Whereas, others are RGB image datasets. **CIFAR-10** consists of 50,000 images for training and 10,000 for testing. There are 5,000 samples for each of the 10 object categories. **CIFAR-100** is constructed in a similar fashion with the same size of the training and testing set. The difference lies in the granularity of the data distribution as 100 classes are categorised (500 images corresponding to each class). The **SVHN** dataset represent 10 digit classes with 73,257 train images and 26,032 test images. Finally, **FashionMNIST** contains training and testing sets of the size 60,000 and 10,000, respectively, with annotations of 10 clothing designs. From an input image resolution perspective, despite FashionMNIST with a 28x28 size, the other datasets have 32x32 scale.

Together with the classification task, we shift the learner's objective to regression. As we tackle the 3D Hand Pose Estimation task, we benchmark our baselines on one of the most challenging, widely been used and first of depth based datasets, **ICVL**[6]. This is composed of 16,004 images for training and 1,600 for testing. The dataset has a single frontal viewpoint and a wide range of articulation and hand positions. The initial resolution is 320x240, but we pre-process by hand centring and scaling to 128x128.

The last benchmark we deployed in the experiment section is the face expression dataset, Radboud Faces Database (**RaFD**)[4]. This is formed of 7,200 training images, 800 for each of the 8 expressions. However, the test set contains only 840 images. Although the initial image dimensions are 256x256x3, for efficiency, we downscale them by a factor of 2. As we consider the entire training set as labelled in this experiment, we generate with StarGAN[1] 57,600 images for the unlabelled set. Similar to the CIFAR-10 evaluation settings, we initially create a randomly distributed subset $\mathbf{D}_S$ of 10,000 images from which we further apply the selection given a budget b of 1,000.

B. Experiments

CIFAR-10 imbalanced dataset In the experimental part, we evaluated quantitatively in a systematic manner the active learning methods over four image classification datasets. Although, before selection, we randomise the unlabelled samples to a subset, the dataset is still relatively balanced to each class distribution. However, this is not commonly the case where there is no prior information re-

lated to the data space. Therefore, we are simulating an imbalanced CIFAR-10 in a quantitative experiment. Beforehand we considered the 50,000 training set as unlabeled, given 5,000 samples for each of the 10 categories. We custom the dataset so that 5 of the 10 classes contain 10 % of their original data (500 samples each). Therefore, the new unlabelled pool is composed of 27,500 images. The experiment architecture and settings are similar to the one on the full scale.

Figure B.1: Quantitative results - CIFAR-10 imbalanced dataset

Figure B.1 shows the progressions of the presented baselines. Our proposed methods, UncertainGCN and CoreGCN, out-stand once again the other model-based selections like VAAL and Learning Loss. UncertainGCN scores 2% more than those methods with 80.05% mean average accuracy at 10,000 labelled samples. Meanwhile, CoreGCN achieves 84.5% top performance together with CoreSet. Thus, the geometric information is more useful in scenarios where the dataset is imbalanced.

Figure B.2: Ablation studies - CIFAR-10 GCN Hyper-parameters tuning

Ablation study - GCN parameter search While varying the architectural parameters of the GCN binary classi er, we encountered a poorer selection with the increase of the Dropout rate from 0.3 to 0.5 or 0.8. However, when changing the size of the hidden units to 256 and 512, the UncertainGCN sampling was not affected on CIFAR-10. This might require further optimisation for different datasets although robustness is being shown.

Figure B.3: CIFAR-10 Learner VGG-11 - 3 selection stages

VGG-11 learner for CIFAR-10 image classi cation for 3 selection stages In Figure B.3, we modi ed the architecture of the learner from CIFAR-10 experiment to VGG-11[5]. Therefore, we analyse how the AL methods are affected in terms of accuracy at the fourth sampling stage. In training the VGG-11 network, we kept the same hyper-parameters. We also had to trace the features after the rst four Max Pooling layers for the Learning Loss baseline. Our proposed methods present robustness to this change whilst GCN settings were left unchanged. Hence, they surpass all state-of-the-arts at this early stage. This also demonstrates how the batch size and the feature representation play an important role in the performances of the other baselines. The most affected baseline in this context is CoreSet.

Hyper-parameters Study Here, we present the analysis of two important hyper-parameters in the objective of the sampler. These are GCN uncertainties margin, s_{margin} and , the labelled vs unlabelled data loss weighing factor. Figure B.5 summarises these studies. From the Figure, we observe that the performance improves when we decrease margin from 0.4 to 0.1. Afterwards, the performance is stable. This shows that our method is stable in the range of an optimal margin. Similarly, in uences the performance. However, the drift in performance is smooth with the change in the value of

Extended qualitative analysis on the AL method In Figure B.4, we extend our qualitative analysis by visualising the initial, the unlabelled and the last selected sam-

Figure B.4: Extended qualitative analysis on labelled/unlabelled images at the last selection stage for CIFAR-10, ICVL and RaFD

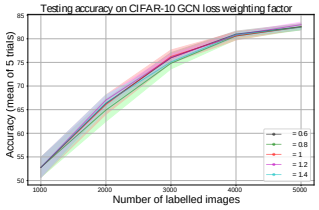

Figure B.5: Hyper-parameter study on UncertainGCN margin (s_{margin}) (left) and labelled vs unlabelled data loss weighing factor, λ (right) (Zoom in the view)

ples from CIFAR-10, ICVL and RaFD. The last selection stage for CIFAR-10 and ICVL is the 10th, while in the synthetic RaFD experiment is the 4th. The seed labelled images are acquired randomly before the first selection stage. The RaFD seed examples are from the entire training set as the AL selection is applied on StarGAN generated images. For all the three benchmarks we evaluated the selected examples with our proposed AL method, UncertainGCN. Although the seed labelled samples for CIFAR-10 are randomly selected, the top query images from the "cat" class consist of difficult examples. On the other hand, the remained unlabelled images present distinguishable features, easy for the learner to predict. These observations have been quantified in the main paper as well. However, in the ICVL dataset case, the selected samples show closer and easier hand articulations compared to the initial labelled set. This is because of the highly complex set that was used as seed examples. The unlabelled images might have a lack of representativeness in the learner's perception after all the 10 sampling stages. Finally, in the RaFD synthetic subsampling process, we can clearly denote the noisy images that were left unlabelled. These present more artefacts than the selected group.

References

[1] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *CVPR*, 2018.

[2] Ian J Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, 2013. 1312.6082v4.

[3] Alex Krizhevsky. Learning multiple layers of features from tiny images. *University of Toronto*, 05 2012.

[4] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, Skyler T Hawk, and AD Van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and emotion*, 2010.

[5] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Network for Large-scale image recognition. In *ICLR*, 2015.

[6] Danhang Tang, Hyung Jin Chang, Alykhan Tejani, and Tae-Kyun Kim. Latent regression forest: Structured estimation of 3d articulated hand posture. In *CVPR*, 2014.

[7] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms, 2017. 1708.07747v2.