Streaming Network Applications for Adverse Weather and Lighting Conditions

Sergey Tarasenko Fumihiko Takahashi
System Development Unit, AI System Dept., Mobility Technologies
Tokyo, Japan
{sergey.tarasenko,fumihiko.takahashi}@mo-t.com

Recently it has been illustrated that Streaming networks (STNets) [1, 2] are capable of recognizing zero noise-corrupted images with moderate accuracy without using special techniques or training data augmentation. STNets are a family of convolutional neural networks, which consist of multiple neural networks (streams). Each stream has different input from other streams. Outputs of all streams are concatenated and fed into a single joint classifier. Each stream takes a unique intensity slice of an input image. In the most recent study [3], it has been shown that STNets outperform a 1-stream simple convnet [3] for various types of noise and image distortions. The study [3] concludes that noise-robust image classification is achieved by the streaming packaging of the simple convents into one multi-stream network equipped with image intensity slices as inputs. The study [3] was conducted using cifar10 corrupted dataset[1] [4], which includes 19 different types of distortions. Here we illustrate results for six types of distortions, i.e., brightness change, contrast change, fog, frost, glass blur, and snow, which are the most relevant to adverse weather conditions, in Fig. 1. In the same study, some preliminary results of STNets application for low light image classification were introduced. In this study, we focus on the issue of low light image classification and introduce some results of the work in progress.

Streaming Networks Architecture STNets are constructed of multiple parallel streams, whose outputs are concatenated and fed into a classifier. During training, params of all streams are tuned independently. Each stream is taking a certain piece of information and its input is different from the inputs of the other streams. In the original studies by Tarasenko and Takahashi [1, 2], each stream had input in the form of the image intensity slice. STNet architecture is presented in Fig. 2. Mathematically, Img^* image representation using orthogonal functions can be described by eq. (1) [5]:

$$Img^* = \sum_{i=1}^{n} \phi_i(x_i, y_i) \tag{1}$$

[1] https://zenodo.org/record/2535967

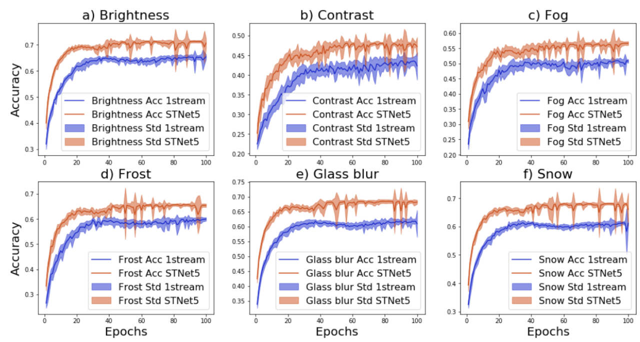

Figure 1. Illustration of robust recognition of corrupted images from Cifar10 Corrupted dataset by 5-stream STNet (STNet5) vs. 1-stream simple convnet.

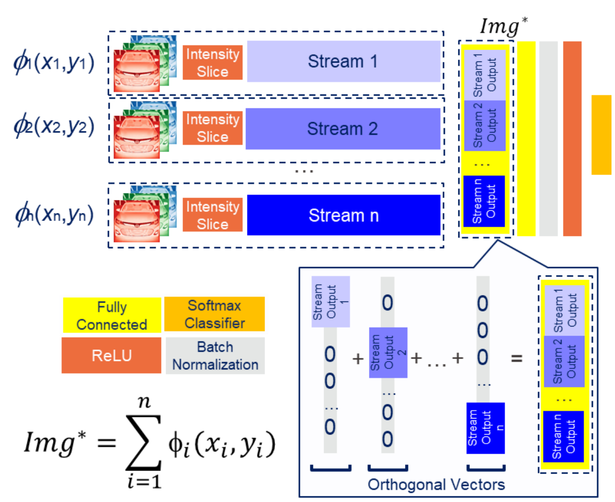

Figure 2. STNet architecture.

where $\phi_i(\cdot)$,i = 1, ..., n, is a function and a pair (x_i, y_i), i = 1, ..., n, represents some part of an original image. Functions $\phi_i(\cdot)$, i=1,...n, correspond to the streams of an STNet, pairs (x_i, y_i), i=1,...n, correspond to intensity slices.

Classification of Low Light Images To test whether STNets can successfully classify low light images after hav-

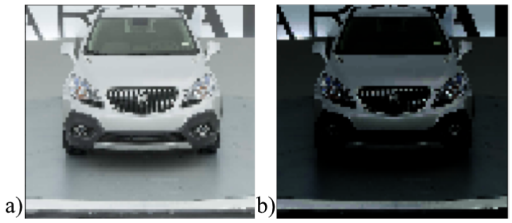

Figure 3. Apply Gamma transform: a) original cropped image; 2) output of Gamma transform with γ=5.6.

ing been trained on normal light images, we employ Carvana[2] dataset. Carvana dataset contains in total 36 car models. We have selected only the front views of all car models. A new dataset consists of 6561 images. For our experiments, we also have resized original Carvana images to 128x191 pixels (height x width). To imitate low light conditions, we have applied Gamma transform with γ=5.6:

$$Img^* = (Img/255)^\gamma * 255 \tag{2}$$

Examples of original lighting condition image and corresponding low light image after Gamma transform are presented in Fig. 3 a) and b), respectively. To illustrate the effect of Gamma transform, we have cropped images to focus only on the area with a car front view.

Throughout experiments, we have used all original 6561 images as the training set (original images). We have also prepared 6561 images after Gamma transform (low light images). To test the networks we used two protocols without (no-aug) and with augmentation (aug). In this context, augmentation implies that randomly selected 50% of low light images are added to original images. In the no-aug case, the training set consisted of original 6561 images and the test set consisted of 6561 low light images. In the aug case, the training set consisted of 9841 (6561 original images + randomly extracted 3280 low light images) and the test set consisted of 3281 low light images remaining after random extraction. Each network was run 5 times.

In total, we have tested eight convnets: VGG16 and VGG19 [6], InceptionResNetV2 [7], MobileNetV2 [8], NASNetMobile and NASNetLarge[3][9], ResNet50 [10], 5-stream STNet (STNet5).

The results are presented in Figs. 4 and 5 for no-aug and aug cases, respectively. These figures introduce best achieved accuracy for each network across 5 runs. Fig. 4 introduces no-aug results and Fig. 5 reports aug results, respectively, vs. model size as number of params, including both trainable and non-trainable params.

Our results indicate that in the no-aug case, STNet5 outperforms all other networks in accuracy while being the smallest network. In the aug case, STNet5 ranks second with ~90% accuracy and 2,464,296 params, while VGG19 ranks first with ~95% accuracy and 73,590,628 params. Thus, VGG19 provides a ~5% accuracy gain while being ~30x bigger.

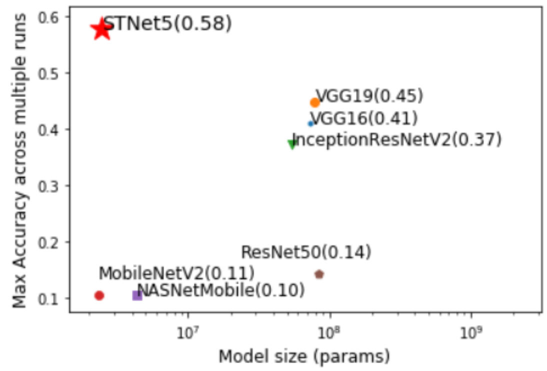

Figure 4. Comparison of convnets' performance for Carvana data in without augmentation during training.

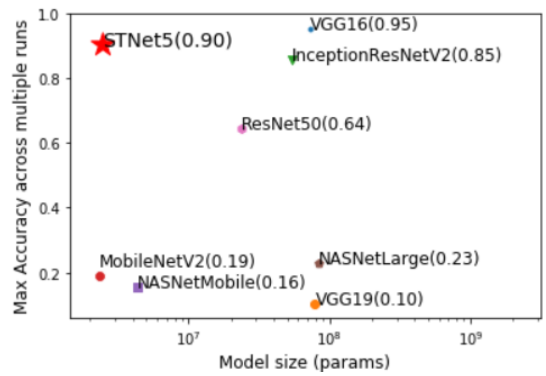

Figure 5. Comparison of convnets' performance for Carvana data with augmentation during training.

Discussion and Conclusion We have illustrated that STNets can effectively deal with various types of image distortion, which can be introduced by adverse weather and lighting conditions. We have tested STNet5 performance against the other eight networks when dealing with low light images. Our results indicate that while being the smallest network STNet5 ranks first in the case of no augmentation training and ranks second in augmentation case, giving up the first place to ~30x bigger VGG16 network. Therefore we conclude that STNets constitute a promising architecture in terms of the balance between efficient usage of network capacity and high accuracy for classification of corrupted images due to adverse weather or lighting conditions.

To continue testing STNets, we plan to extend our experiment for more neural networks and include new techniques and datasets designated to the adverse weather (fog [11] and rain [12]), and lighting conditions (nighttime [13]).

[2]https://www.kaggle.com/c/carvana-image-masking-challenge
[3]only used in augmentation case

References

[1] S. Tarasenko and F. Takahashi (2019) Streaming Networks: Enable A Robust Classification of Noise-Corrupted Images, arXiv:1910.11107. 1

[2] S. Tarasenko and F. Takahashi (2020) Streaming Networks: Increase Noise Robustness and Filter Diversity via Hard-wired and Input-induced Sparsity. arXiv:2004.03334 1

[3] S. Tarasenko and F. Takahashi (2020) Applications of the Streaming Networks, *In the Proceeding of the 34th Annual Conference of Japanese Society for Artificial Intelligence* (in press). 1

[4] D. Hendrycks and T. G. Dietterich (2018) Benchmarking Neural Network Robustness to Common Corruptions and Surface Variations, arXiv:1807.01697. 1

[5] B. A. Olshausen and D. J. Field (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature*, 381, p. 607-609. 1

[6] K. Simonyan and A. Zisserman (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition, Arxiv, abs:1409.1556. 2

[7] Ch. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi (2016) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, *AAAI*. 2

[8] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov and L.-Ch. Chen (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 4510-4520 2

[9] B. Zoph and V. Vasudevan, J. Shlens and Q. V. Le (2017) Learning Transferable Architectures for Scalable Image Recognition, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 8697-8710. 2

[10] K. He, X. Zhang, S. Ren and J. Sun (2015) Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 16)*, p. 770-778. 2

[11] D. Dai, Ch. Sakaridis, S. Hecker and L. V. Gool (2019) Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding,*International Journal of Computer Vision*, p. 1-23. 2

[12] Sh. S. Halder, J.-F. Lalonde and R. de Charette (2019) Physics-Based Rendering for Improving Robustness to Rain, *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, p. 10202-10211. 2

[13] Ch. Sakaridis, D. Dai and L. V. Gool (2019) Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation,*2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, p. 7373-7382.

2