FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

STUDENTSKÁ VĚDECKÁ KONFERENCE 2022

Reducing detection network overfitting with Domain Adaptation Transformations

Jiří Vyskočil[1]

1 Introduction

A known issue of overfitting or underfitting of the neural network can be caused by incorrect hyperparameter settings or insufficient data for the task. While hyperparameters are affected by various influences, e.g., a network architecture, a limited amount of data always results in limited network capabilities of generalization. In other words, it is difficult to train a network to recognize, for example, a rotated object that the network has never seen in this way. Therefore, it is important to add augmentations to the training to see "the new" objects.

The basic types of augmentations are brightness, contrast, flip, or blur. Hence all of these augmentations bring a new distribution into the training. Besides, shifting input space to the mean of each channel also rapidly helps the network to learn faster. This is standardly used in popular APIs, e.g., Detectron2. Therefore, there arises the question, why apply augmentations to cover the wider distribution of input space instead of reducing the shift between two distributions by changing the style of the images? In this work, Domain Adaptation Transformations, originally designed for mapping from synthetic data (source) to real data (target), are applied before the image is passed through the network. Differently to the other approaches, source and target images are from the same dataset. One target image with a suitable illumination is selected from which all input images will be recolored. This is finally evaluated on data from the similar and different distributions in order to observe overfitting and generalization.

2 Datasets

Multiple-Object Tracking datasets, which contain suitable data for the network generalization analysis, were split into training, validation, and testing data. The MOT17 dataset was selected for training using 95% of its images. The performance was validated on the remaining 5% of the images from the end sequences of this dataset. The MOT20 dataset is used to observe the generalization of the network due to a new distribution of video sequences.

3 Domain Adaptation Transformations

Two Domain Adaptation Transformations were used for the study: Fourier Domain Adaptation (FDA) introduced by Yang et al. (2020) and Histogram Matching (HM) from article by Neumann et al. (2005). These transformations transfer style of the target image to the source. HM manipulates the pixels to match histograms of both reference and source images. While FDA transfer the style using Fast Fourier Transform (FFT) by replacing low frequency values of the source amplitude with the values from the target.

[1] PhD. student, University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics, e-mail: vyskocj@kky.zcu.cz

4 Ablation study and results

Faster R-CNN architecture with ResNet-50 backbone and Feature Pyramid Network feature extractor was selected and trained using Detectron2 API and publicly available checkpoints from the COCO dataset. The impact of the transformations is evaluated using standard COCO metrics for Object Detection. All models are trained with the same settings: 30 training epochs, SGD optimizer with momentum of 0.9, smooth L_1 loss, base learning rate of 0.0025 with decay factor of 0.5 at 20th and 25th epoch and warm-up for the 1st epoch. The models differ only in transformation used to the input image. The baseline model has no transformation of image colors, while the others use FDA or HM before the image is passed through the network.

Transform Type	AP		AP_{50}		AP_{75}		AR	
	MOT17	MOT20	MOT17	MOT20	MOT17	MOT20	MOT17	MOT20
None (baseline)	68.3	13.7	87.3	29.5	73.5	9.5	80.0	17.2
FDA ($\beta = 0.01$)	67.5	13.2	86.8	27.6	72.9	9.9	79.9	16.5
FDA ($\beta = 0.001$)	67.9	13.2	86.9	27.9	72.9	9.7	80.1	16.5
FDA ($\beta = 0.0001$)	67.9	12.9	87.1	27.0	73.3	10.0	79.9	15.9
HM (ratio = [1.0 only])	67.8	18.4	86.3	40.1	72.9	13.0	80.3	23.3
HM (ratio = [0.5 to 1.0])	68.0	**20.4**	**87.8**	**45.1**	**73.6**	**13.5**	80.2	**26.4**
HM (ratio = [0.5 only])	**68.4**	17.1	87.5	38.0	73.4	11.8	**80.5**	22.1

Table 1: Comparison of Domain Adaptation Transforms - Fourier Domain Adaptation (FDA) and Histogram Matching (HM) - on MOT17 (validation) and MOT20 (test) datasets. Hyper-parameters used are denoted in brackets at each transform.

Results of the performed study are summarized in Table 1. The results show an reduced network performance while using FDA in both datasets, and HM has no obvious impact on data of a similar distribution. Significant improvement of HM was achieved in all cases on MOT20 dataset, whose images are very different from the training ones. The best results were obtained by choosing the interval from which the images were randomly blended during the training. This proves that Domain Adaptive Transformations can seriously help the network to detect objects on unseen data distribution, e.g., containing night scenes or more objects to detect.

Acknowledgement

Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic. The work has been supported by the grant of the University of West Bohemia, project No. SGS-2022-017.

References

Yang, Y., & Soatto, S. (2020) FDA: Fourier Domain Adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.*

Neumann, L., & Neumann, A. (2005) Color style transfer techniques using hue, lightness and saturation histogram matching. In *Computational Aesthetics in Graphics, Visualization and Imaging.*