Self-ensembling for visual domain adaptation

Geoff French – g.french@uea.ac.uk

Colour Lab (Finlayson Lab)

University of East Anglia, Norwich, UK

Image montages from http://www.image-net.org

Thanks to

My supervisory team: Prof. G. Finlayson, Dr. M. Mackiewicz

Competition organisers and all participants

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Described in more detail in our ICLR 2018 submission

"Self-Ensembling for Visual Domain Adaptation"

https://arxiv.org/abs/1706.05208 (v2)

Model

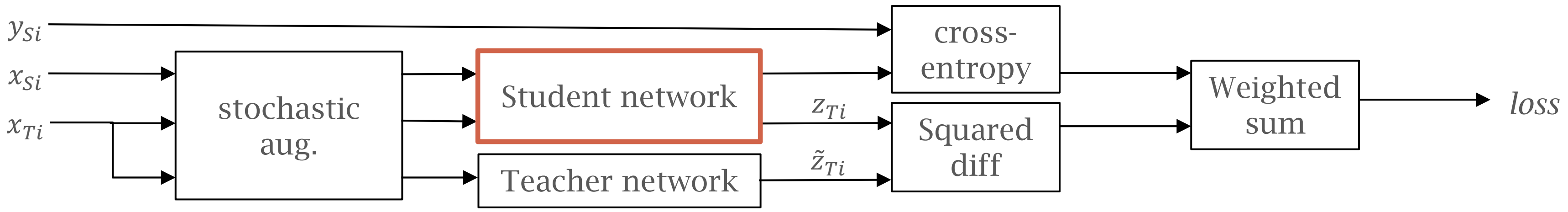
Self-ensembling developed for semi-supervised learning in [Laine17]

Further developed in [Tarvainen17] (mean teacher model)

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Mean-teacher model

Standard classifier DNN

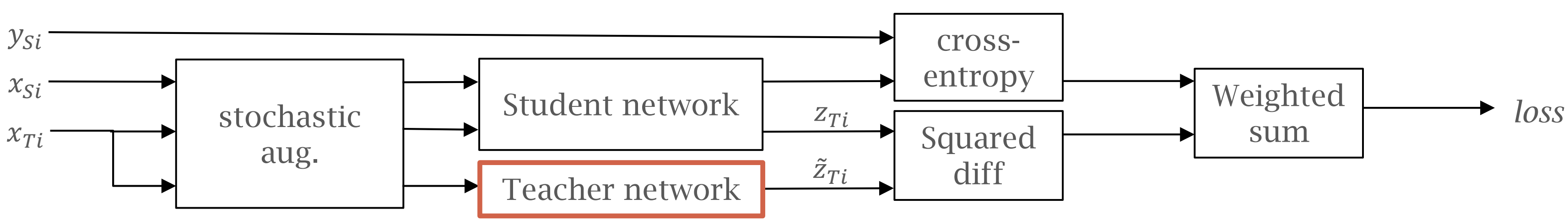


https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Mean-teacher model

Weights of teacher network are exponential moving average of student network

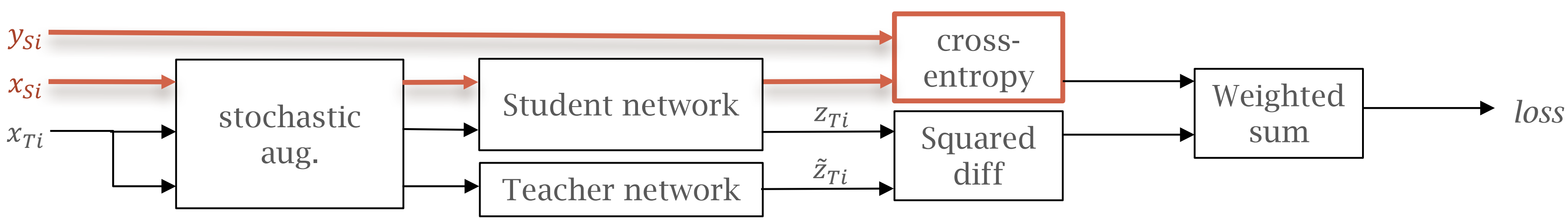


https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Source domain sample:

Traditional supervised cross-entropy loss (with data augmentation)

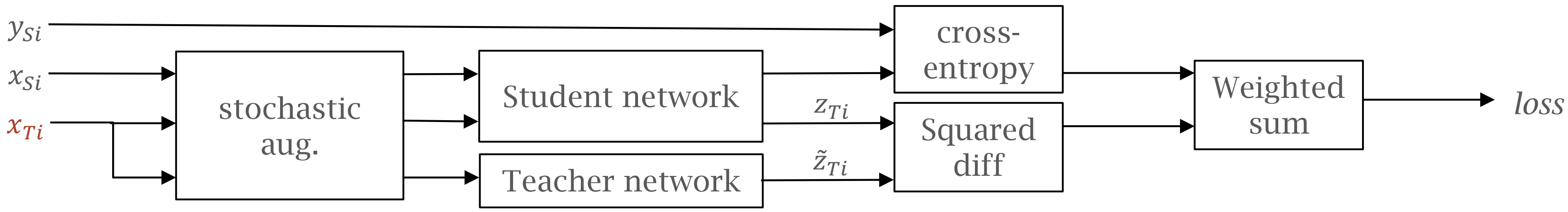


https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Target domain sample:

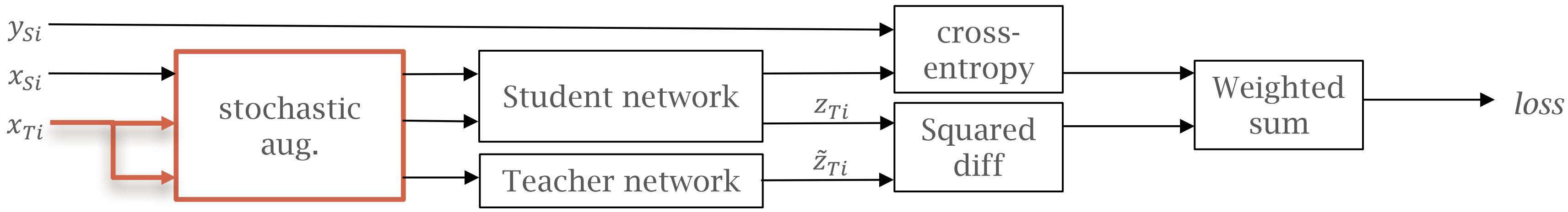
one sample



https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Target domain sample:

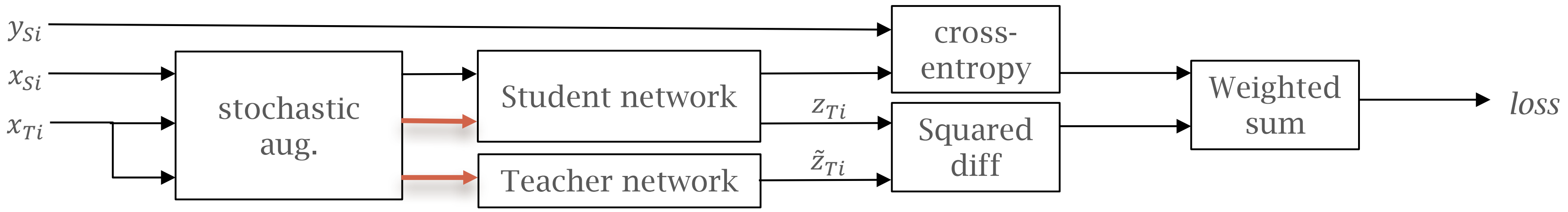
augment twice, differently each time (gaussian noise, translation, flip)



https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Target domain sample:

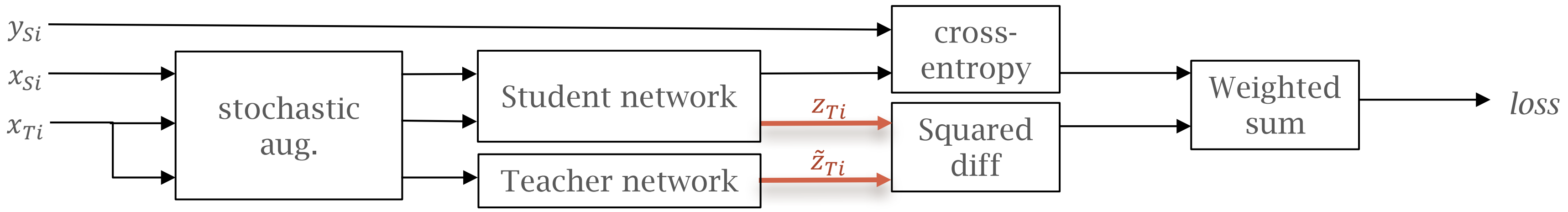
One path through student network
Second through teacher
(different dropout)



https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Target domain sample:

Result: two predicted probability vectors

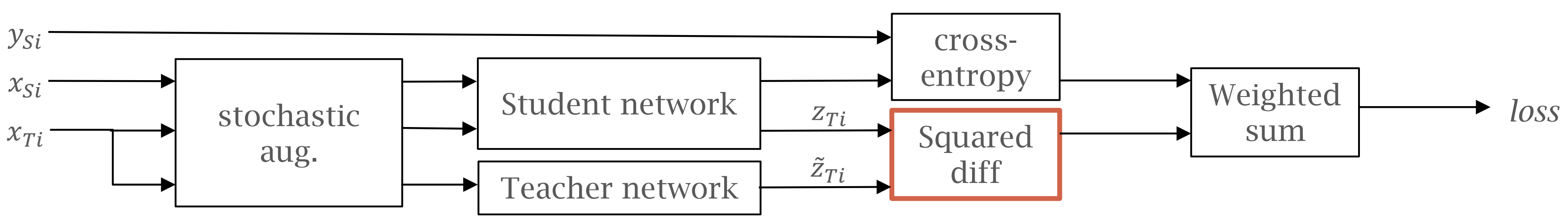


https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Target domain sample:

Self-ensembling loss: train network to learn to make them the same (squared difference)



https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Self-ensembling performs label propagation over unsupervised samples

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Model so far may handle simple domain adaptation tasks…

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Our adaptations for domain adaptation

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Separate source and target batches

Per training iteration, process source and target mini-batches separately

Each gets its own batch norm stats, bit like AdaBN [Li16]

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Confidence thresholding

If confidence of teacher predictions <96.8%, mask self-ensembling loss for that sample to 0

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

MORE Data augmentation

VisDA model

Random crops, rotation, scale, h-flip
Intensity/brightness scaling, colour offset, colour rotation, desaturation

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

MORE Data augmentation

Our small image benchmarks:

$$\begin{bmatrix} 1+\mathcal{N}(0,0.1) & \mathcal{N}(0,0.1) \\ \mathcal{N}(0,0.1) & 1+\mathcal{N}(0,0.1) \end{bmatrix}$$

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Class balancing

Binary-cross-entropy loss between target domain predictions (averaged over sample dimension) and uniform probability vector

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Class balancing

Otherwise in unbalanced datasets one class is re-inforced more than the others

Classifier separates source domain from target and assigns all target domain samples to most populous class

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Works with randomly initialised nets

e.g. for small image benchmarks

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Works with pre-trained nets

e.g. the ResNet 152 we used for VisDA ☺

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

VisDA-17 Results

https://arxiv.org/abs/1706.05208

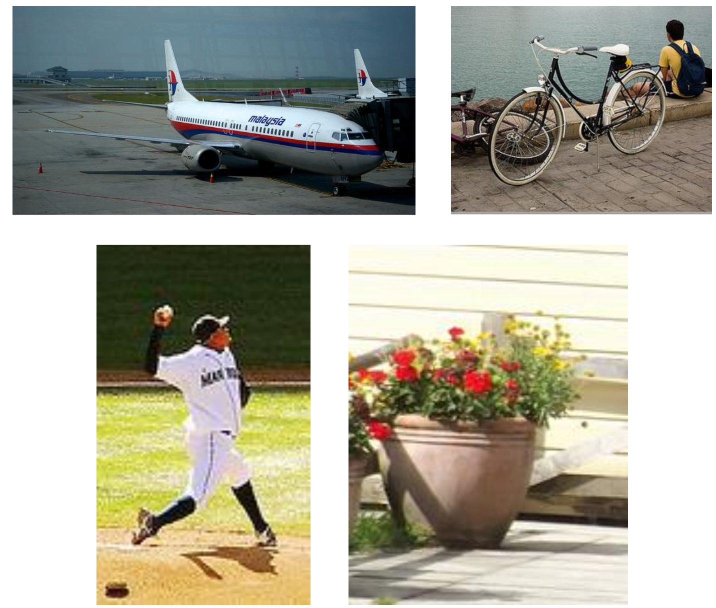
Self-Ensembling for Visual Domain Adaptation

Images from VisDA-17

Training set

Labeled

Validation set



Unlabeled

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Model

Fine-tuned ResNet-152

Remove classification layer (after global pooling)

Replace with two fully-connected layers

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Notes

Test set augmentation

Predictions were computed by augmenting each test sample 16x and averaging predictions

Gained 1-2% MCA on validation set

Notes

5 network ensemble

Predictions of 5 independent training runs were averaged

Gained us ~0.5% MCA on test set

VisDA-17

VALIDATION	Acc	TEST	Acc
Resnet-50	*82.8*	Resnet-50	
Resnet-152		Resnet-152	

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

VisDA-17

VALIDATION	Acc	TEST	Acc
Resnet-50	*82.8*	Resnet-50	
Resnet-152	**85.3***	Resnet-152	

* Not on leaderboard

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

VisDA-17

VALIDATION	Acc	TEST	Acc
Resnet-50	*82.8*	Resnet-50	~80
Resnet-152	**85.3***	Resnet-152	

* Not on leaderboard

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

VisDA-17

VALIDATION	Acc	TEST	Acc
Resnet-50	*82.8*	Resnet-50	~80
Resnet-152	**85.3***	Resnet-152	***92.8***

* Not on leaderboard

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

VisDA-17

VALIDATION	Acc	TEST	Acc
Resnet-50	*82.8*	Resnet-50	~80
Resnet-152	**85.3***	Resnet-152	***92.8***

	Plane	Bike	Bus	Car	Horse	Knife	
Val	96.3	87.9	84.7	55.7	95.9	95.2	
Test	96.9	92.4	92.0	97.2	95.2	98.8	

	MCycle	Person	Plant	Skbrd	Train	Truck	MEAN
Val	88.6	77.4	93.3	92.8	87.5	38.2	**82.8**
Test	86.3	75.3	97.7	93.3	94.5	93.3	**92.8**

* Not on leaderboard

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

Validation set

Lots of confusion between car and truck

Much less so on test set

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Small image results

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

MNIST ⟷ USPS

MNIST

USPS

Model	USPS → MNIST	MNIST → USPS
Sup. on SRC	91.97	96.25
SBADA-GAN [Russo17]	97.60	95.04
OURS	**99.54**	**98.26**
Sup. On TGT	99.62	97.83

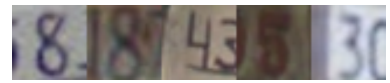
https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Syn-digits → SVHN

Syn-digits

SVHN



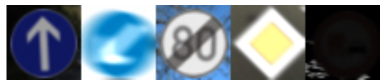
Model	Syn-digits → SVHN
Sup. on SRC	86.96
ATT [Saito17]	93.1
OURS	**96.00**
Sup. On TGT	95.55

https://arxiv.org/abs/1706.05208

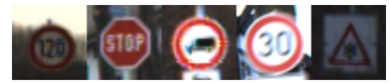
Self-Ensembling for Visual Domain Adaptation

Syn-signs → GTSRB

Syn-signs



GTSRB



Model	Syn-signs → GTSRB
Sup. on SRC	96.72
ATT [Saito17]	96.2
OURS	**98.32**
Sup. On TGT	98.54

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

SVHN (greyscale*) → MNIST

SVHN (grey)

MNIST

Model	SVHN → MNIST
Sup. on SRC	73.00
ATT [Saito17]	76.14
OURS	**99.22**
Sup. On TGT	99.66

* [Ghiffary16]

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

MNIST → SVHN (greyscale)

MNIST

SVHN (grey)

Model	MNIST → SVHN
Sup. on SRC	28.78
SBADA-GAN [Russo17]	**61.08**
OURS	41.98
Sup. on TGT	96.68

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

MNIST → SVHN (greyscale)

MNIST (aug)

SVHN (grey)

Model	MNIST → SVHN
Sup. on SRC (aug)	64.82
SBADA-GAN [Russo17]	61.08
OURS	**96.6**
Sup. on TGT (aug)	97.3

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Conclusions

Our approach has produced good results

It won VisDA ☺

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

Promising avenue for domain adaptation: two components…

https://arxiv.org/abs/1706.05208

Self-Ensembling for Visual Domain Adaptation

STEP 1. Align source and target distributions

Pre-trained net, data augmentation, …

Prior work in the field (e.g. CORAL, AdaBN) does this!

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

STEP 2. refine correspondence

Self-ensembling is well suited to this

https://arxiv.org/abs/1706.05208
Self-Ensembling for Visual Domain Adaptation

THANK YOU!

References

[Ghiffary16] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. "Deep reconstruction-classification networks for unsupervised domain adaptation.“ *ECCV* 2017.

https://arxiv.org/abs/1706.05208

[Laine17] Samuli Laine and Timo Aila . "Temporal Ensembling for Semi-Supervised Learning.“ *ICLR* 2017.

[Li16] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. "Revisiting batch normalization for practical domain adaptation." 2016.

https://arxiv.org/abs/1706.05208

[Saito17] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada . "Asymmetric Tri-training for Unsupervised Domain Adaptation“ 2017.

https://arxiv.org/abs/1706.05208

[Russo17] Paolo Russo, Fabio Maria Carlucci, Tatiana Tommasi, and Barbara Caputo
. "From source to target and back: symmetric bi-directional adaptive GAN.“ 2017.

https://arxiv.org/abs/1706.05208

[Tarvainen17] Antti Tarvainen and Harri Valpola. "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results.“ 2017.

https://arxiv.org/abs/1706.05208