SEBASTIAN SCHERER, ROBIN SCHÖN, RAINER LIENHART

PSEUDO-LABEL NOISE SUPPRESSION TECHNIQUES FOR SEMI-SUPERVISED SEMANTIC SEGMENTATION

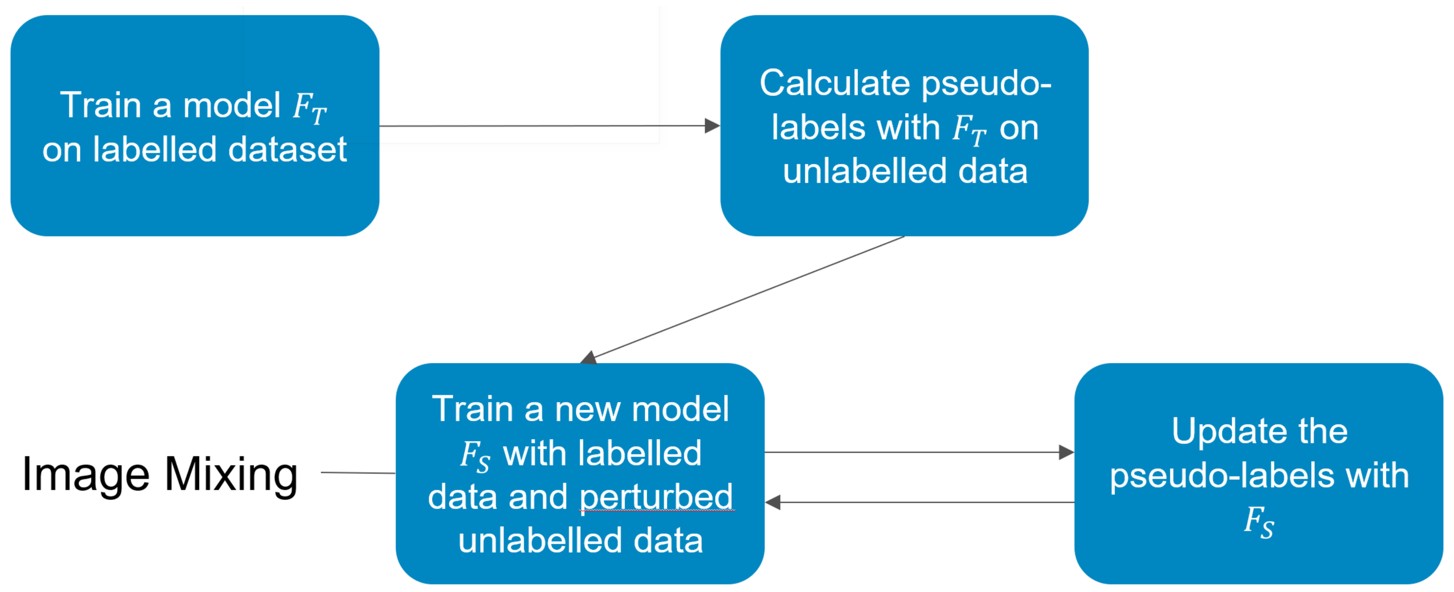
Universität Augsburg University

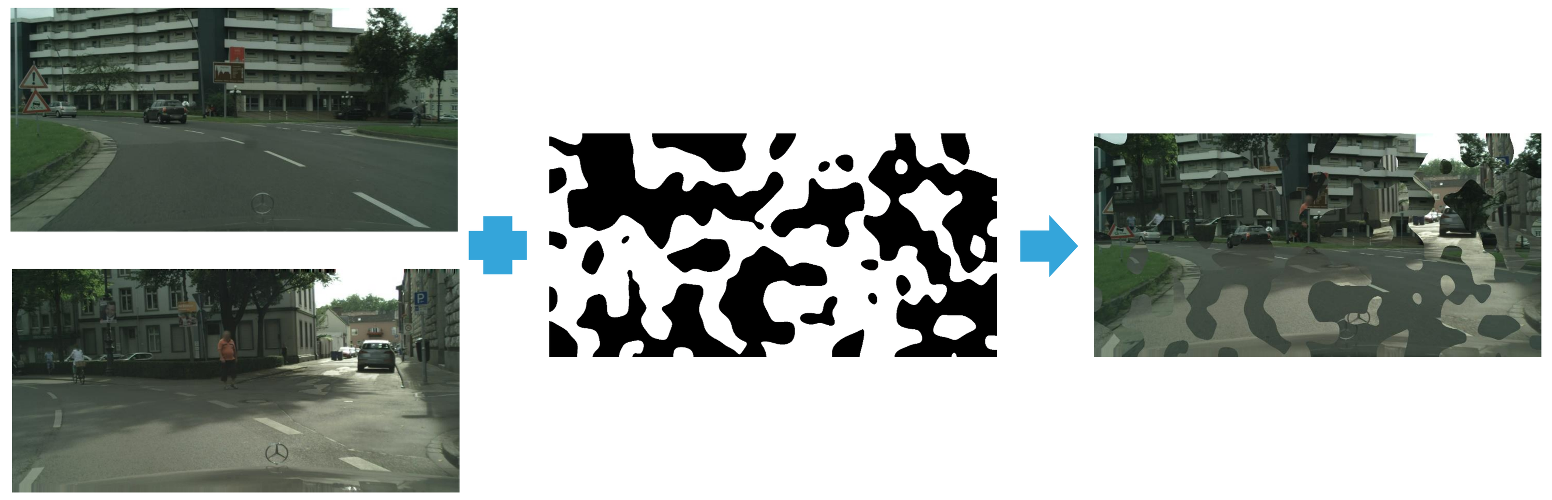
INTRODUCTION

- Semi-supervised learning (SSL) reduce the need for large labelled datasets by incorporating unlabelled data into the training. This is particularly interesting for semantic segmentation, where labelling data is very costly and time-consuming.
- Current SSL approaches use an supervised trained model to generate predictions for unlabelled images, called pseudo-labels, which are subsequently used for training a new model from scratch.
- Since the predictions do not come from an error-free neural network, they are naturally full of errors. However, training with partially incorrect labels often reduce the final model performance. Thus, it is crucial to manage errors/noise of pseudo-labels wisely.
- In this work, we use three mechanisms to control pseudo-label noise and errors:
- (1) We construct a base framework by mixing images with cow-patterns to reduce the negative impact of wrong pseudo-labels.
- (2) We propose a simple and effective loss weighting scheme for pseudo-labels defined by the feedback of the model trained on these pseudo-labels.
- (3) We study the common practice to ignore pseudo-labels with low confidence during training and empirically analyse the influence and effect of pseudo-labels with different confidence ranges and the contribution of pseudo-label filtering to the achievable performance gains.

METHOD

- We use the iterative pseudo-label refinement method and first train an initial segmentation model on the labelled dataset.
- The model is used as teacher F_T to generate pseudo-labels for unlabelled images.
- A new student model F_S is trained on the labelled and unlabeled dataset, while latter used the pseudo-labels from F_T.
- Cycle process where F_S becomes the F_T at next iteration.
- To prevent the model from memorizing wrong pseudo-labels, we utilize image mixing as perturbation when training with pseudo-labels [1].





Pseudo Label Filtering

- Remove pseudo-labels with low confidence from training and assume that a low confidence correlates with wrong pseudo-labels
- A higher confidence is equated with a higher certainty that the pseudo-label is correct.

Pseudo Label Weighting

- **Observation:** The confidence of correct label predictions grow faster than the confidences of wrong label predictions over the course of model training.
- We propose a weighting scheme with the aim to assign smaller weights to possibly wrong pseudo-labels and therefore reduce the contribution of noisy pseudo-labels at training.
- Further adaption of the Symmetric Cross Entropy Loss as more robust loss function to outliers.
- With $p^i = F_S(\bar{x}_u^i)$ as prediction of the model on the perturbed image, $\hat{y}_u^i = argmax(F_T(x_u^i))$ as pseudo-labels and α, β as balancing coefficients the loss is defined as:

$$l^i = w^i \odot (\alpha l_{CE}(p^i, \hat{y}_u^i) + \beta l_{CE}(\hat{y}_u^i, p^i))$$

- The dynamic weights is defined by the softmax output of F_S:

$$w^i = P(\hat{y}_u^i) \ with \ P = F_S(x_u^i)$$

ABLATION STUDY

- Ablation Study on three different datasets: Cityscapes [2] (C), Mapillary [3] (M) and PASCAL Voc 2012 [4] (P), with 15, 100 and 183 labelled images respectively.
- **ST**: Self Training, **CM**: CowMask, **PLF**: Pseudo-Label Filtering, **PLW**: Pseudo-Label Weighting, **SCE**: Symmetric Cross-Entropy Loss.

Model	CM	ST	ST_{CM}	PLF	PLW	PLW_{SCE}	Iter.	C(15)	M(100)	P(183)
DL3+								53.0	50.7	57.0
DL3+	✓							52.1	50.2	56.2
DL3+		✓						56.2	53.9	66.3
DL3+			✓					60.3	56.1	68.1
DL3+			✓	✓				60.9	56.8	67.8
DL3+			✓		✓			61.6	57.0	68.9
DL3+			✓			✓		62.4	57.8	69.5
DL3+			✓	✓		✓		63.0	58.0	69.3
DL3+			✓	✓		✓	✓	66.5	60.1	-

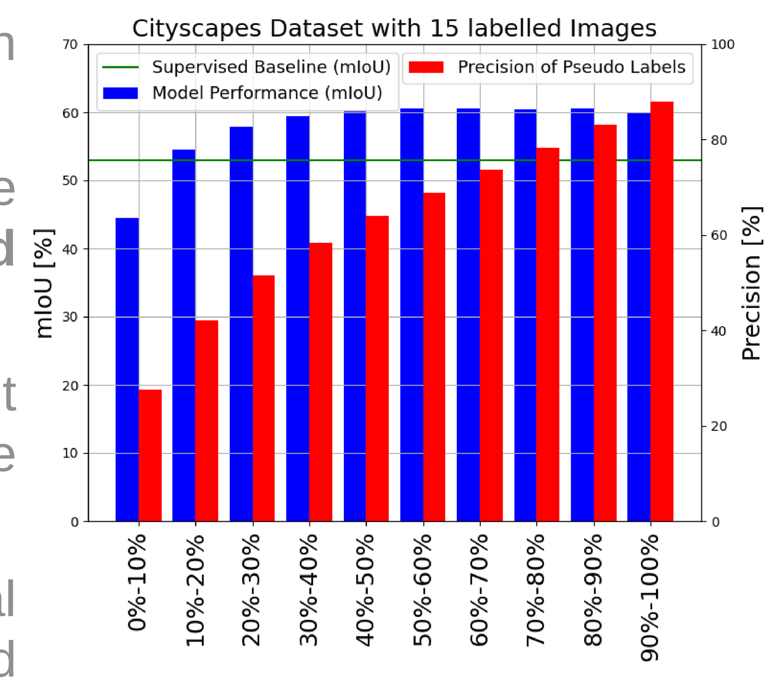
EXPERIMENTS

- Cityscapes Dataset [2]: 2975 training and 500 validation images.
- We use different fraction of ratios between labelled and unlabelled data from the training set.

Method	Model	$\frac{1}{200}$ (15)	$\frac{1}{20}$ (148)	$\frac{1}{16}$ (186)	$\frac{1}{8}$ (372)	$\frac{1}{4}$ (744)	Full
SupOnly (ours)	DL3+	53.0	64.1	64.8	69.0	73.3	78.7
CutMix* [11]	DL3+	-	-	67.1	71.8	76.4	-
Yuan et al. [28]	DL3+	-	-	-	74.1	77.8	78.7
U2PL [24]	DL3+	-	-	74.9	76.5	78.5	-
PS-MT [17]	DL3+	-	-	-	77.1	78.4	-
CPS [5]	DL3+	-	-	74.7	77.6	**79.1**	80.4
Ours	DL3+	**66.5**	**74.4**	**75.7**	**78.0**	78.7	-

Pseudo Label Filtering Experiments

- Analysis of pseudo-labels with different confidence.
- Clear correlation between confidence and pseudo-label correctness (**red bars**).
- Incorporating different confident samples for SSL training, we observe a weaker correlation (**blue bars**).
- We observe only marginal improvements with confidence-based pseudo-label filtering.



Qualitative Results

- Example Prediction on unseen data with a DeepLabV3 model trained semi-supervised using only 15 labelled images.

HUMAN POSE ESTIMATION

- LSP [5] dataset with 2000 images and 14 joints, split into 1600 training and 400 validation images.
- Simple Unet architecture with ResNet18 backbone used.
- Metric: Percenage of correct keypoints (PCK) with 0.2 ratio.
- We observe similar findings as in semantic segmentation.

PL: Training with Pseudo Labels, **CM**: CowMask, **PLF**: Pseudo-Label Filtering, **PLW**: Pseudo-Label Weighting

Method	$\frac{1}{16}$ (100)	$\frac{1}{4}$ (400)	Full (1600)
Sup. only	62.3	73.5	82.6
PL	65.9	76.4	-
+ CowMask	68.2	78.2	-
+ PLW	69.6	78.8	-
+ PLF	69.4	78.5	-
+ iter	71.3	79.4	-

References

[1] French et al., Milking cowmask for semi-supervised image classification, 2020
[2] Cordts et al., The cityscapes dataset for semantic urban scene understanding, 2016
[3] Neuhold et al., The mapillary vistas dataset for semantic understanding of street scenes, 2017
[4] Everingham et al., The pascal visual object classes (voc) challenge, 2010
[5] Sam Johnson and Mark Everingham, Clustered pose and nonlinear appearance models for human pose estimation, 2010