Grid Loss: Detecting Occluded Faces

Michael Opitz Georg Waltner Georg Poier Horst Possegger Horst Bischof
Graz University of Technology, Institute for Computer Graphics and Vision, Austria

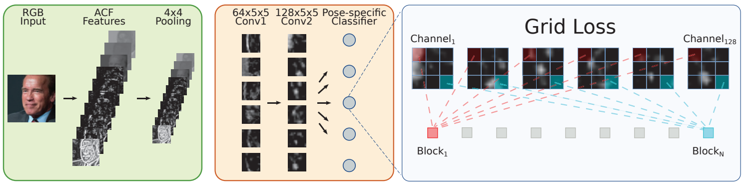
Motivation and Contribution

- Recent benchmark evaluations (*e.g.* FDDB [2]) show that **standard CNN detectors fail on occluded faces**.
- **State-of-the-art** approaches **rely on** large **pre-trained** ImageNet **models**, *e.g.* [1], and on large datasets with **face-attribute annotations** to re-rank object proposals [4].
- We show that **simple CNN based detectors** trained from scratch can outperform these methods, if they **handle occlusions properly**.
- To this end, we train a part-based CNN detector with our **grid loss function**.

Standard Loss Grid Loss

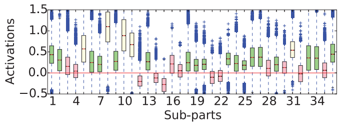
- **Supplemental material** is available **online.**

Overview

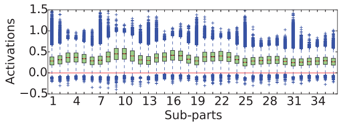


- 2 convolution layers on top of Aggregate Channel Features.
- Linear **pose-specific classifiers** on top of the last convolution layer.
- At test time: **fully convolutional** detection over an image pyramid.
- **Regressor** to refine location of detected faces.
- To tackle occlusions we look at spatially non-overlapping **blocks** on the **last convolution layer**.
- **Grid loss optimizes a loss on each of these blocks separately.**

Loss Function



Standard Loss



Grid Loss

- **Median response** of a CNN detection template on the **positive** training set is **negative with standard loss functions.**
- We encourage a CNN to **make sub-regions** of the detection template **discriminative**:
  - Divide the last convolution layer $\boldsymbol{f}$ into **blocks** $\boldsymbol{f}_i$.
  - Optimize loss on blocks separately to train **part detectors** $\boldsymbol{w}_i$.
  - Share weights with a **regular layer** $\boldsymbol{w}$.

$$l(\boldsymbol{\theta}) = \max(0, 1 - y \cdot (\boldsymbol{w}^\top \boldsymbol{f} + b)) + \lambda \cdot \sum_{i=1}^{N} \max(0, m - y \cdot (\boldsymbol{w}_i{}^\top \boldsymbol{f}_i + b_i))$$

Robustness to Occlusions

- **Parts are discriminative alone.**
- If subset of parts fail (due to occlusion), detection can recover.

Method	COFW-HO	COFW-LO
Grid Loss	**0.979**	**0.998**
Standard Loss	0.909	0.982

True Positive Rate on COFW Heavily Occluded (COFW-HO) and Less Occluded (LO) subsets with our loss vs standard loss.

Diversity of Learned Features

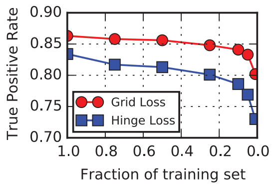
- **More diverse features** compared to standard loss.
- Prevents learning only a small subset of prominent features, *e.g.* eye.
- Encourages learning **discriminative features for all sub-regions**.

Method	Correlation
Grid Loss	**225.96**
Standard Loss	22500.25

Correlation on feature maps

Generalization Ability

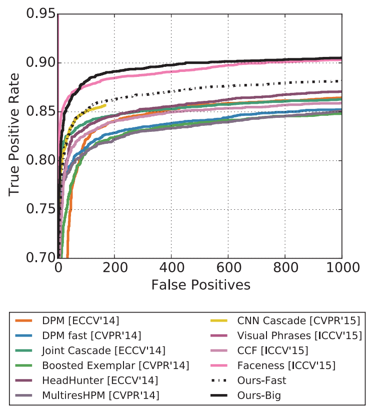
- **Better generalization** ability, due to larger diversity of features.
- On **smaller training set sub-sets** the **performance gap** between Grid Loss and standard loss functions **increases**.



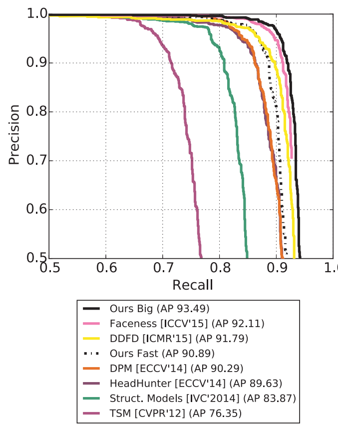
Comparison of true positive rate with detectors trained on a subset of the dataset

Benchmark Results

- We achieve **state-of-the art performance** on several datasets.



Evaluation on FDDB [2]



Evaluation on PASCAL Faces [3]

References and Acknowledgments

[1] S. S. Farfade, M. Saberian, and L.-J. Li. Multi-view Face Detection Using Deep Convolutional Neural Networks. In *Proc. ICMR*, 2015.

[2] V. Jain and E. Learned-Miller. FDDB: A Benchmark for Face Detection in Unconstrained Settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.

[3] J. Yan, X. Zhang, Z. Lei, and S. Z. Li. Face Detection by Structural Models. *IVC*, 32(10):790 – 799, 2014.

[4] S. Yang, P. Luo, C.-C. Loy, and X. Tang. From Facial Parts Responses to Face Detection: A Deep Learning Approach. In *Proc. ICCV*, 2015.

[5] X. Zhu and D. Ramanan. Face Detection, Pose Estimation and Landmark Estimation in the Wild. In *Proc. CVPR*, 2012.

This work was supported by the Austrian Research Promotion Agency (FFG) project DIANGO (840824).