

## Introduction

- Multi task deep learning: learning multiple tasks using shared representations in a supervisory environment
- Existing multi-task learning methodologies rely on splitting the network at a particular layer depending on the task. They do not generalize well across various tasks.

**Objective:** Explore split architecture and cross-stitch on facial landmark dataset. Regularize weight to share representation by introducing a mutual weight regularization loss.

- Cross stitch unit combines representations from both networks through learnable parameters.
- Similarity in tasks implies relatedness in weights of the networks, hence weight regularization might be helpful.

## Problem Formulation

Split Architecture: Facial landmark detection using auxiliary tasks of attribute detection. The last fully connected layer is split. Joint loss function

$$\arg \min_{W^T \{W^a\}_{a \in A}} \sum_{i=1}^N (\ell^r, f(x_i; W^r)) + \sum_{i=1}^N \sum_{a \in A} \lambda^a \ell^a(y_i^a; f(x_i, W^a))$$

Cross-stitch network: two networks are cross-stitched to share representations using learnable cross-stitch unit

$$\begin{pmatrix} \tilde{x}_A^{ij} \\ \tilde{x}_B^{ij} \end{pmatrix} = \begin{pmatrix} \alpha_{AA} & \alpha_{AB} \\ \alpha_{BA} & \alpha_{BB} \end{pmatrix} \begin{pmatrix} x_A^{ij} \\ x_B^{ij} \end{pmatrix}$$

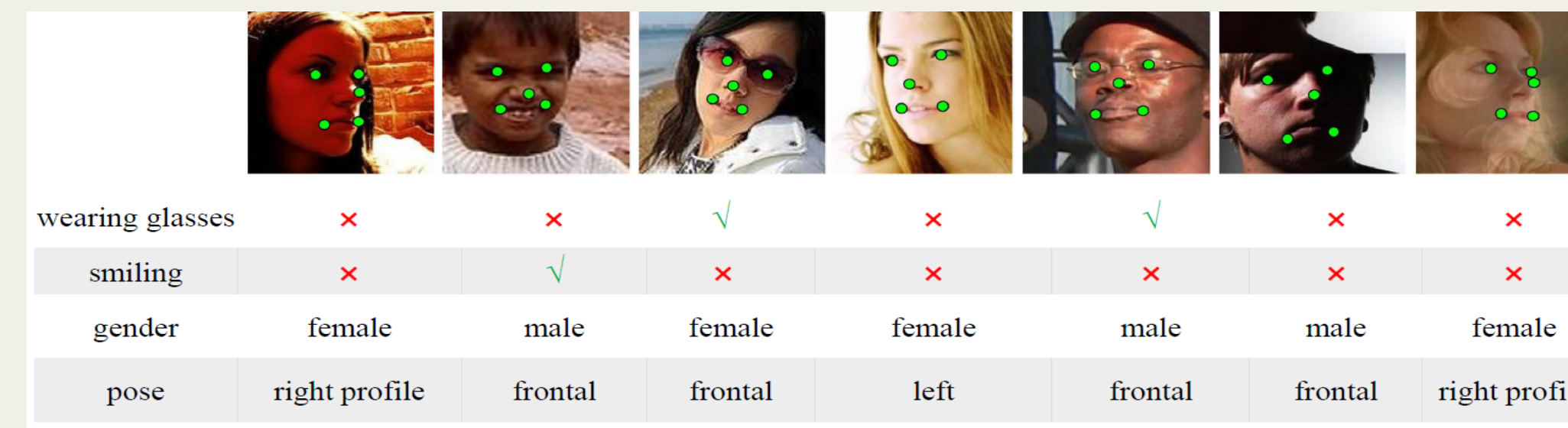
Proposed Weight Regularization: Introducing loss in the weights of two multi-tasking streams to relate the weights

$$L_w = \lambda_w \sum_{j \in \Omega} r_w(\theta_j^{task1}, \theta_j^{task2})$$

$$L_{total} = L_{task1} + L_{task2} + L_w$$

## Dataset

### Facial Landmark Detection + Attribute Prediction



Original data has 5 facial landmark (left eye, right eye, nose, left mouth corner and right mouth corner) and 4 attributes.

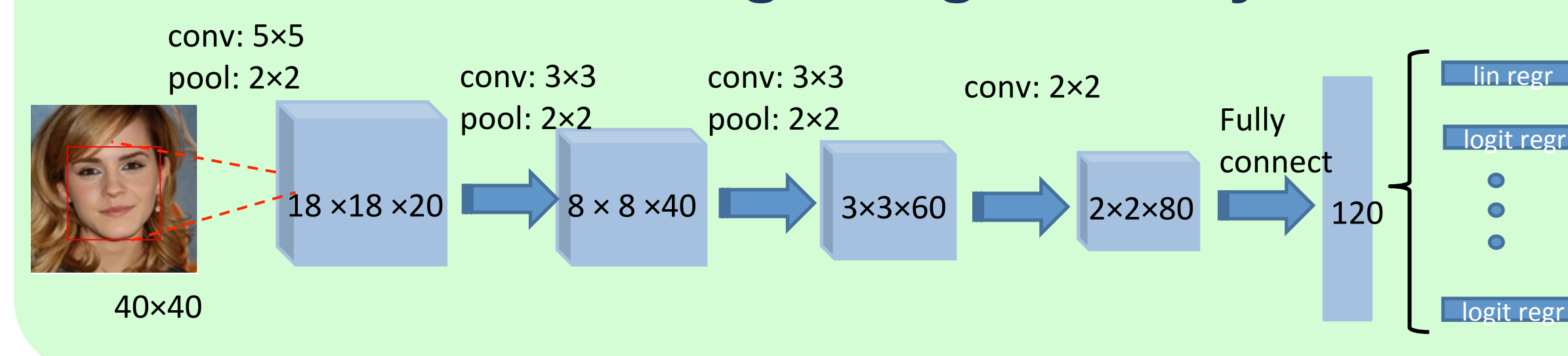
Smiling	Gender	Pose	Glass
S/NS	M/F	0, ±30, ±60	W/NW

Pre-processing:  
Bounding box regression (40×40)

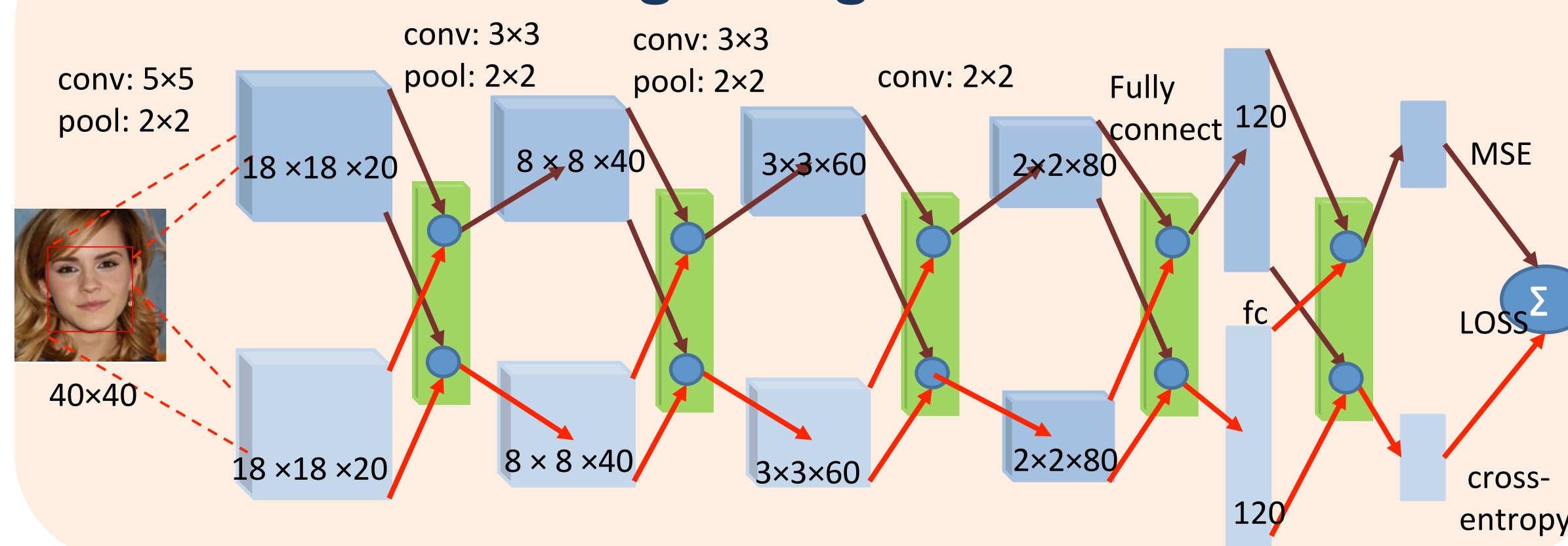


## Methodology

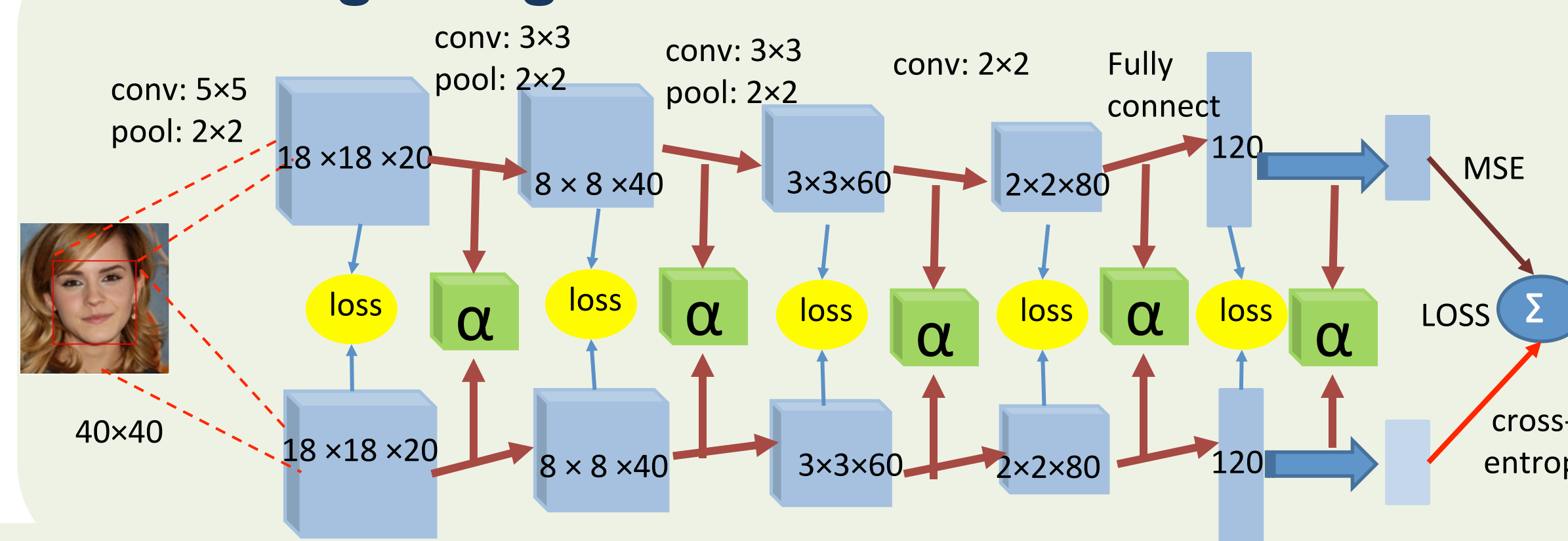
### Multi-task learning using auxiliary tasks



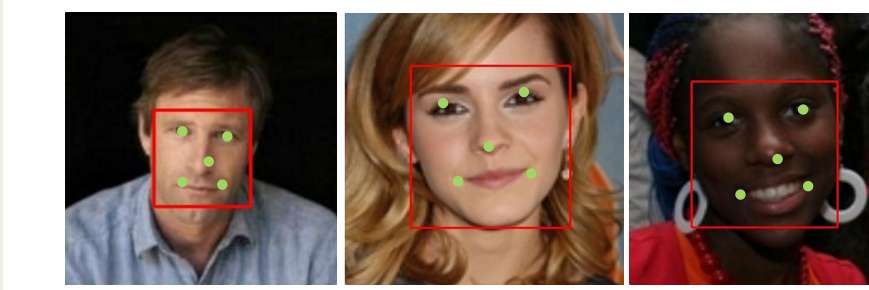
### Multi-task Learning using cross-stitch network



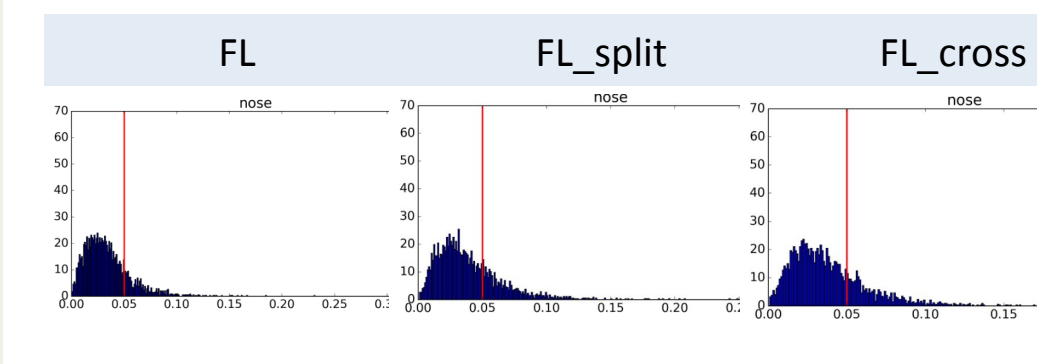
### Weight regularization in cross-stitch



## Result



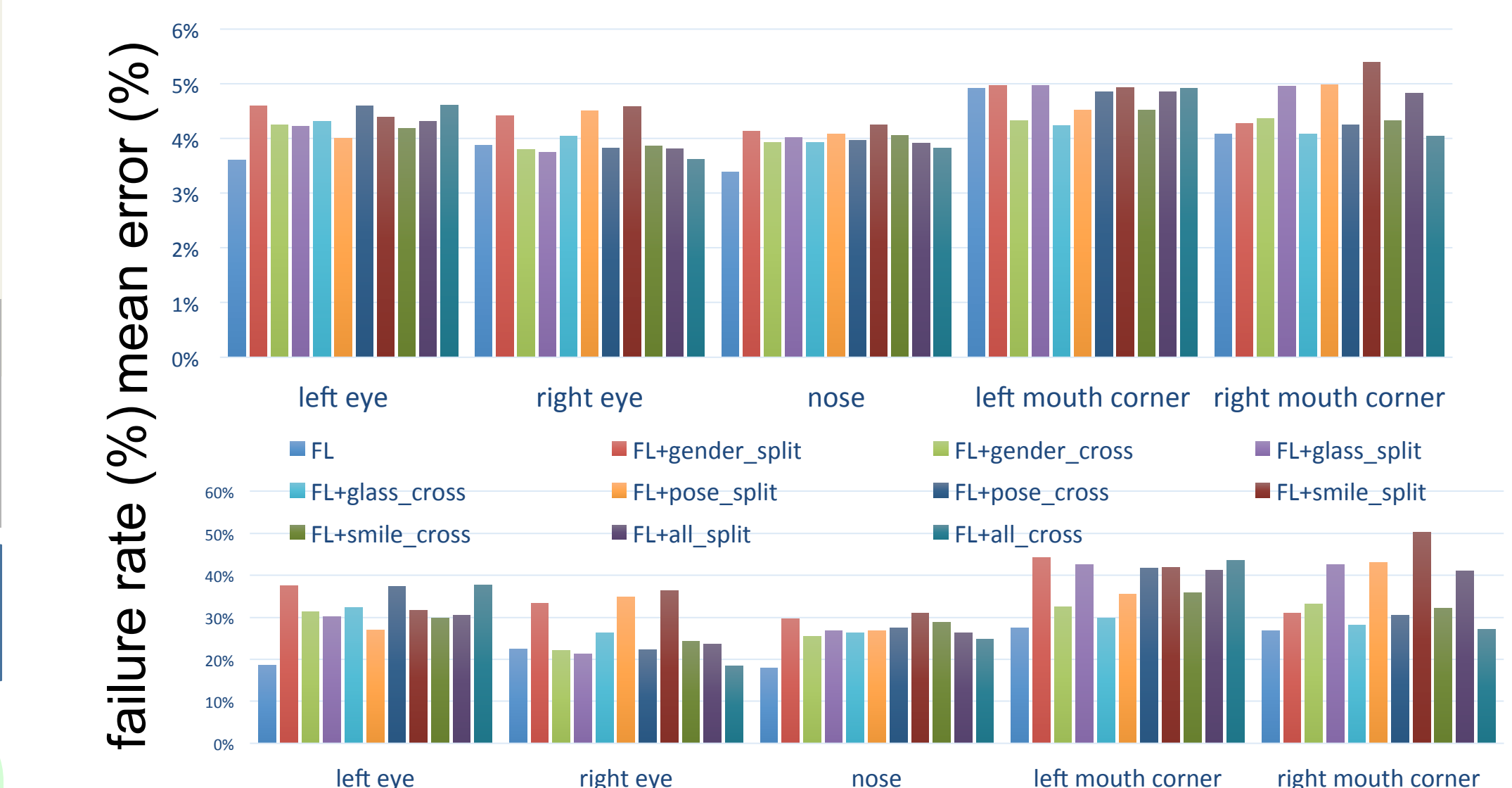
Facial landmark : left eye, right eye, nose, left mouth corner, and right mouth corner



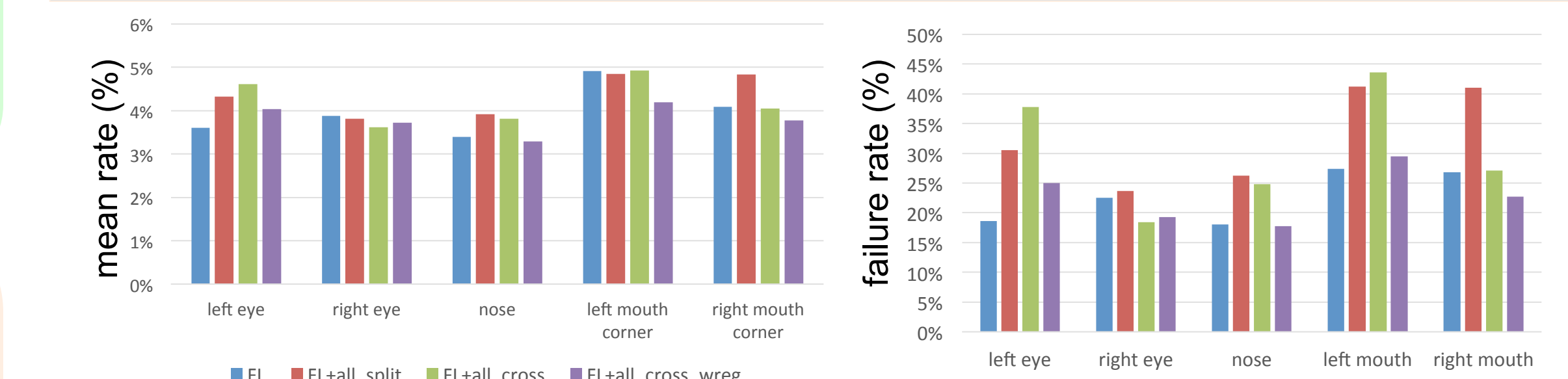
Mean error: L2 norm of difference in prediction and ground truth normalized w.r.t. bounding box width.

Failure rate: mean error larger than 5%

Comparison of different model variants of FLD: mean error rate and failure rate



Proposed weight regularization in cross-stitch: mean error rate and failure rate



## Conclusion & Future Work

- Auxiliary tasks degrade the convergence of the main task (FLD) by overfitting the training set. An early stopping mechanism can be applied. (when validation error below some threshold)
- Cross stitch is applied in all the layers for better generalization, however depending on the tasks it might not converge well. Extensive tuning of parameters is necessary to get the intended result
- Weight regularization with cross stitch gives better result

### References

Misra, Ishan, et al. "Cross-stitch Networks for Multi-task Learning." *arXiv preprint arXiv:1604.03539* (2016).  
Rozantsev et al. "Beyond Sharing Weights for Deep Domain Adaptation." *arXiv preprint arXiv:1603.06432*(2016).  
Zhang, Zhanpeng, et al. "Facial landmark detection by deep multi-task learning." *European Conference on Computer Vision*. Springer International Publishing, 2014.