

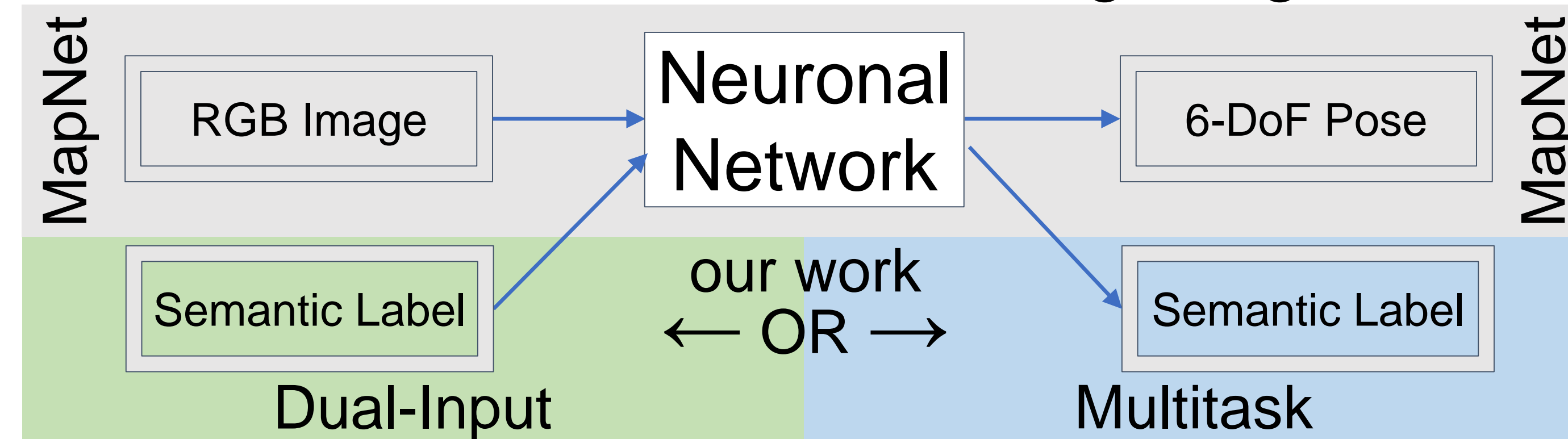
Absolute camera localization

Advanced Deep Learning for Computer Vision

Leonhard Feiner, Alexander Ziller

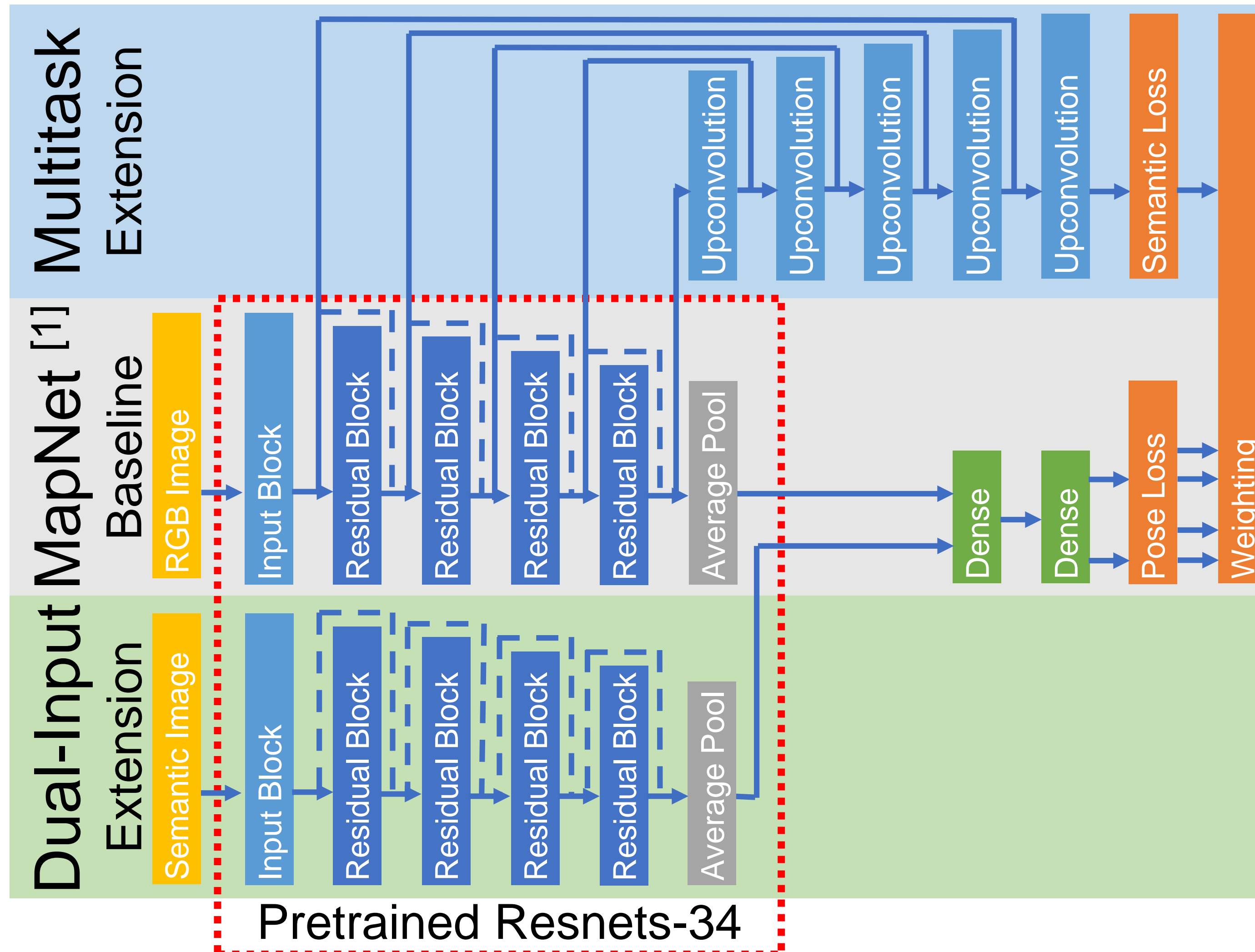


Calculation of Camera Poses using Images



Goal: Improve localization accuracy

Architectures



Weighting of Losses:

$$\mathcal{L}^{combined} = \mathcal{L}^{weighted}_{trans,abs} + \mathcal{L}^{weighted}_{rot,abs} + \mathcal{L}^{weighted}_{trans,rel} + \mathcal{L}^{weighted}_{rot,rel} (+ \mathcal{L}^{weighted}_{semantics})$$

Uncertainty Weighting [3]:

$$\mathcal{L}_X^{weighted} = \frac{1}{2\sigma_X^2} \mathcal{L}_X + \log(\sigma_X)$$

$$\mathcal{L}_{semantics}^{weighted} = \frac{1}{\sigma_X^2} \mathcal{L}_X + \log(\sigma_X)$$

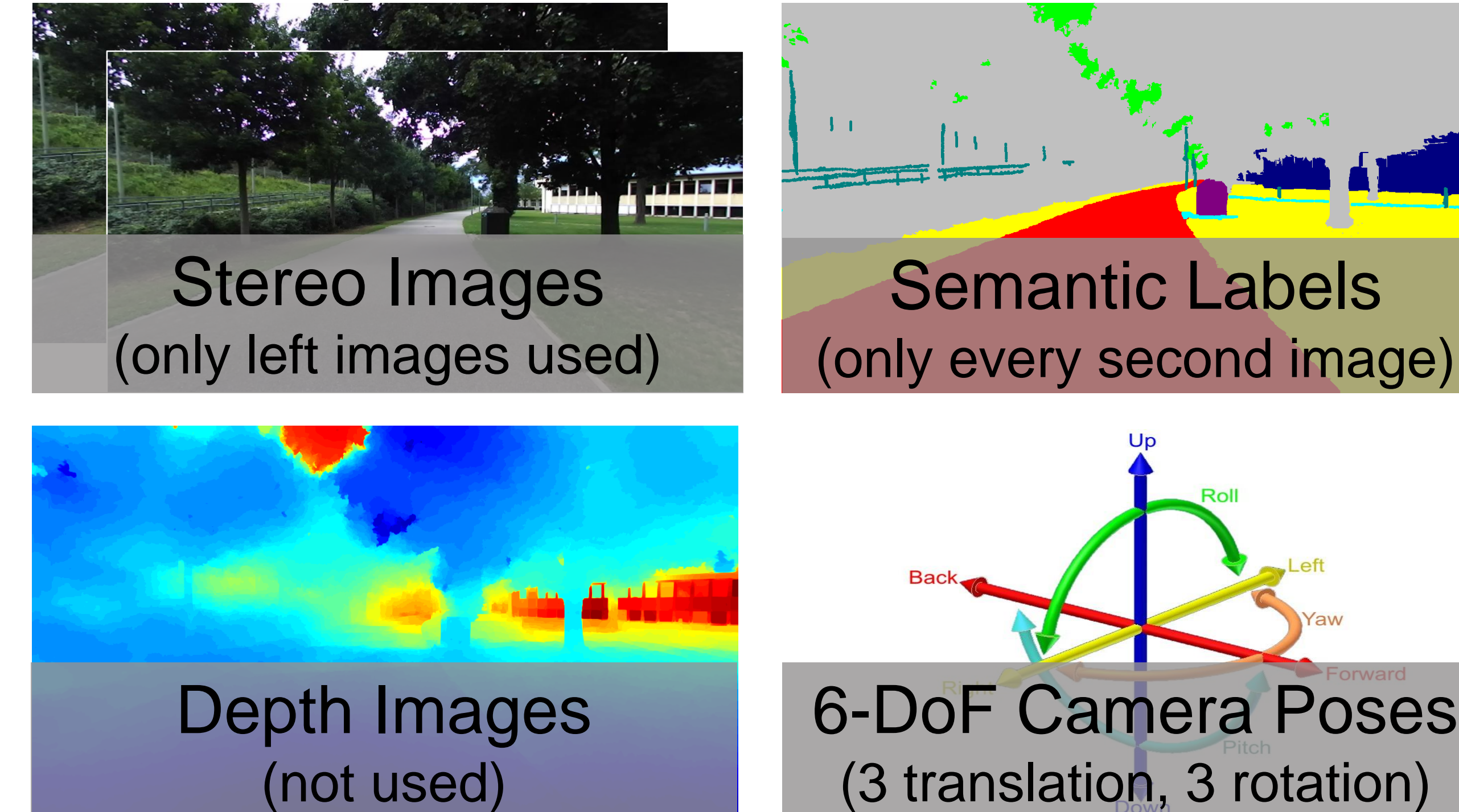
Geometric Weighting:

Used in MapNet, Dual-Input

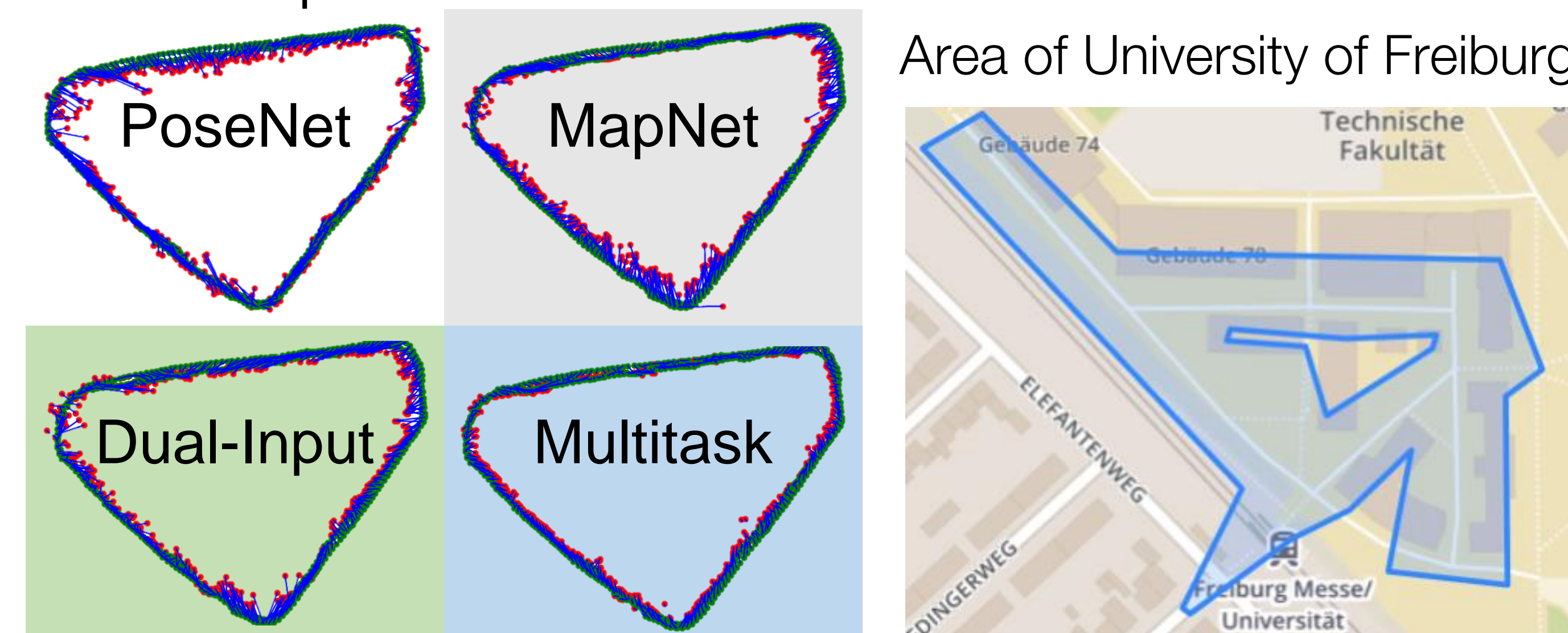
$$\mathcal{L}_X^{weighted} = e^{-\alpha_X} \mathcal{L}_X + \alpha_X$$

$X \in \{(trans, abs); (rot, abs); (trans, rel); (rot, rel)\}$
 α and $\log(\sigma)$ are learnable weights

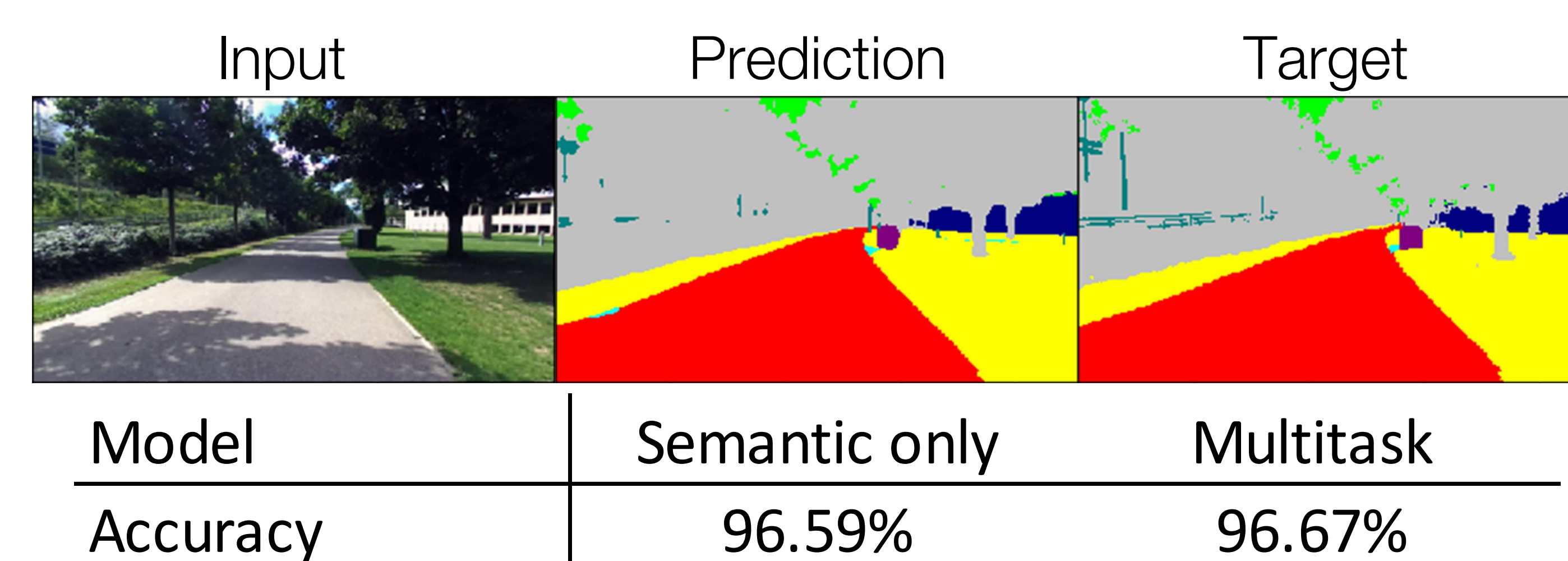
Dataset DeepLoc [2]



Pose Graphs on Validation Set



Semantic Prediction on Validation Set

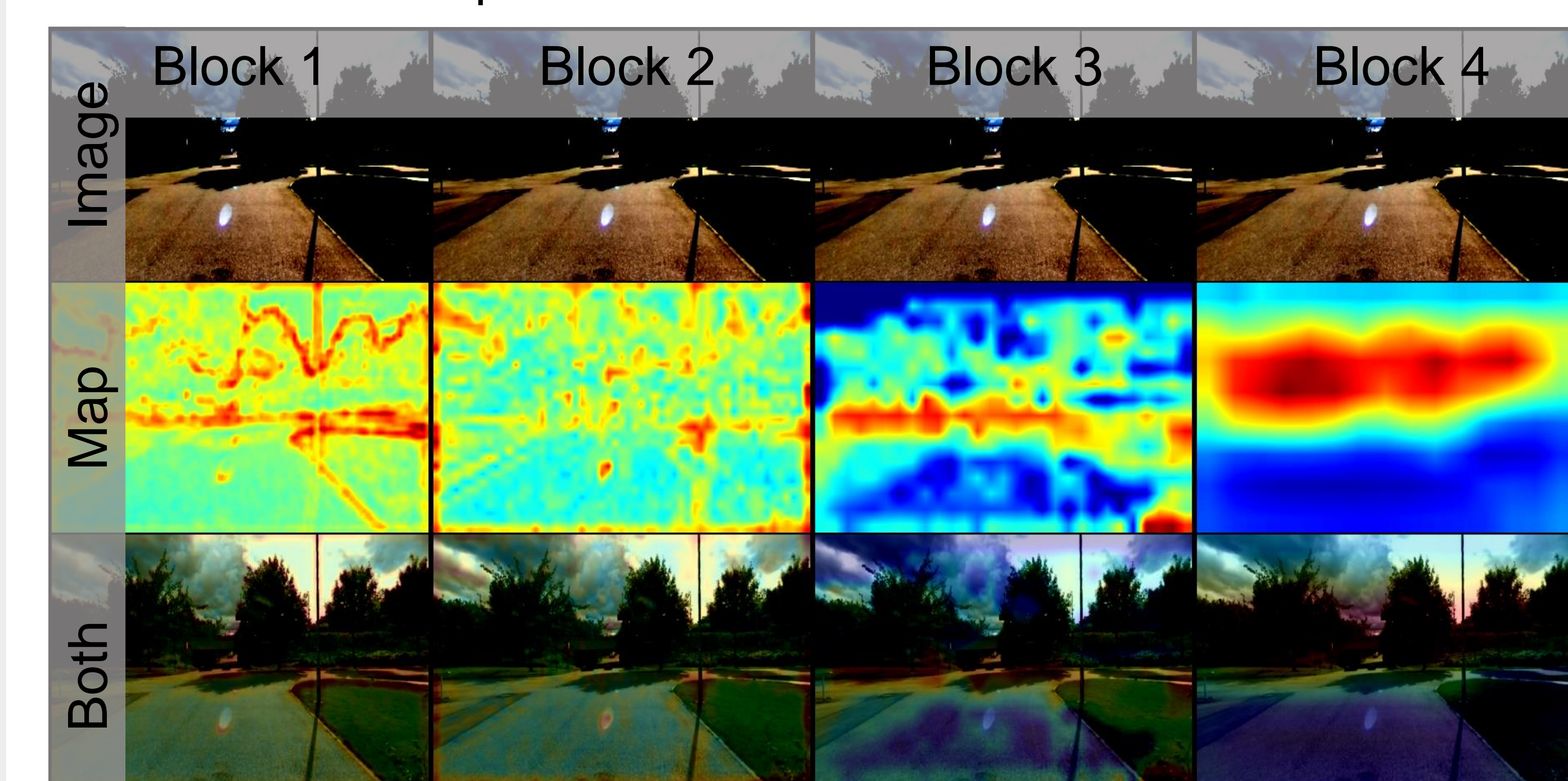


Pose Accuracy on Validation Set

Model	Source	Translation		Rotation	
		Median	Mean	Median	Mean
Posenet	Our	3.64m	4.44m	3.13°	4.27°
Posenet	[2]	2.42m		3.66°	
MapNet	Our	3.42m	3.91m	3.23°	4.48°
Dual-Input	Our	3.26m	3.54m	2.82°	4.32°
Multitask	Our	1.07m	1.31m	2.96°	4.32°
VLocNet	[2]	0.68m		3.43°	
VLocNet++	[2]	0.37m		1.93°	

Initialization influences results significantly

Activation Maps of Residual Blocks [4]



Conclusion

- Semantic information improves pose regression
- Useful as additional semantic input or output
- Multitask learns further insights about environment

[1] Brahmbhatt, Samarth, et al. "Mapnet: Geometry-aware learning of maps for camera localization." *arXiv preprint arXiv:1712.03342* (2017).

[2] Radwan, Noha, Abhinav Valada, and Wolfram Burgard. "Vlocnet++: Deep multitask learning for semantic visual localization and odometry." *arXiv preprint arXiv:1804.08366* (2018).

[3] Kendall, Alex, Yarin Gal, and Roberto Cipolla. "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics." *arXiv preprint arXiv:1705.07115* 3 (2017).

[4] Chattopadhyay, Aditya, et al. "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks." *2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.*