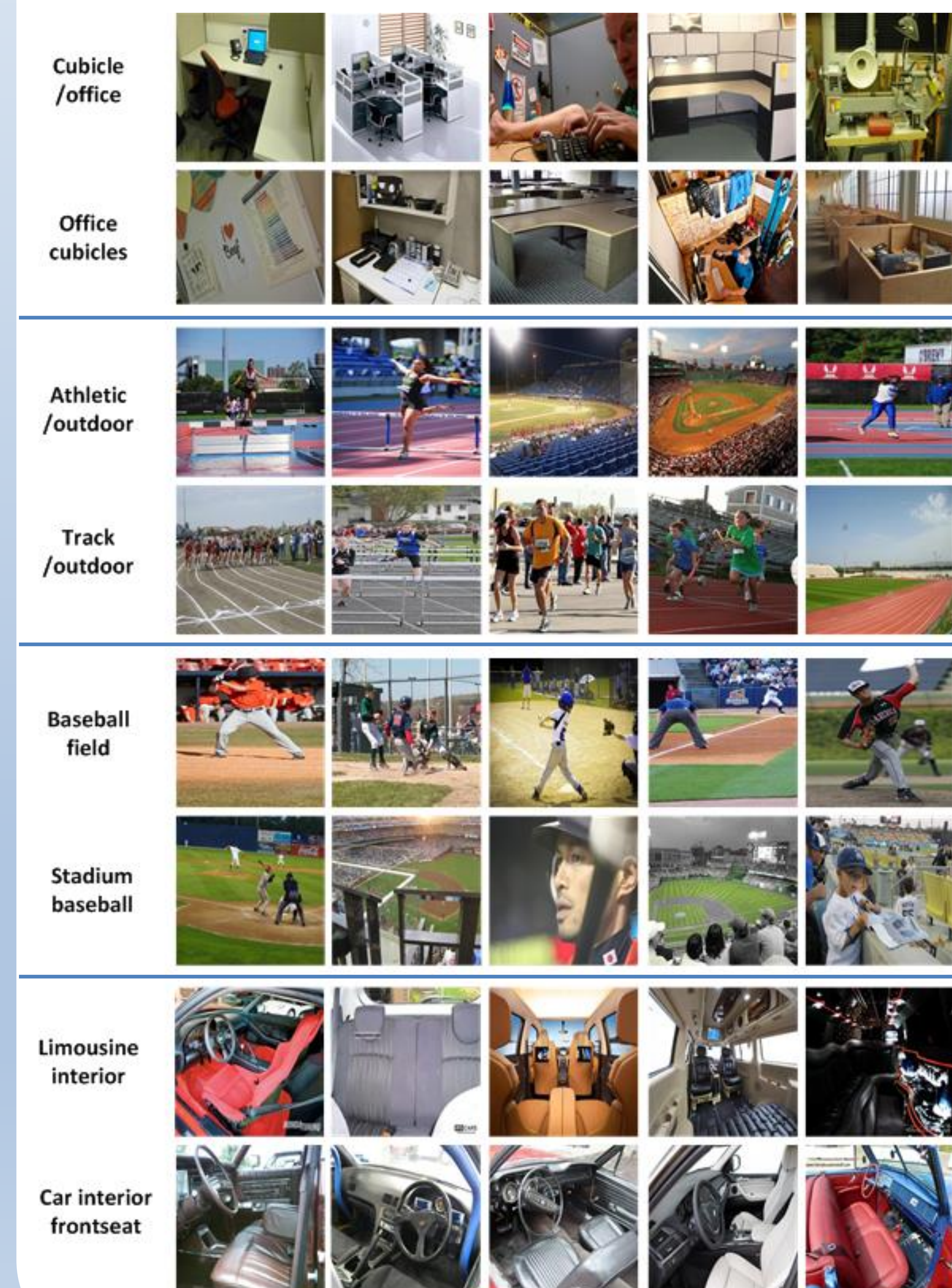


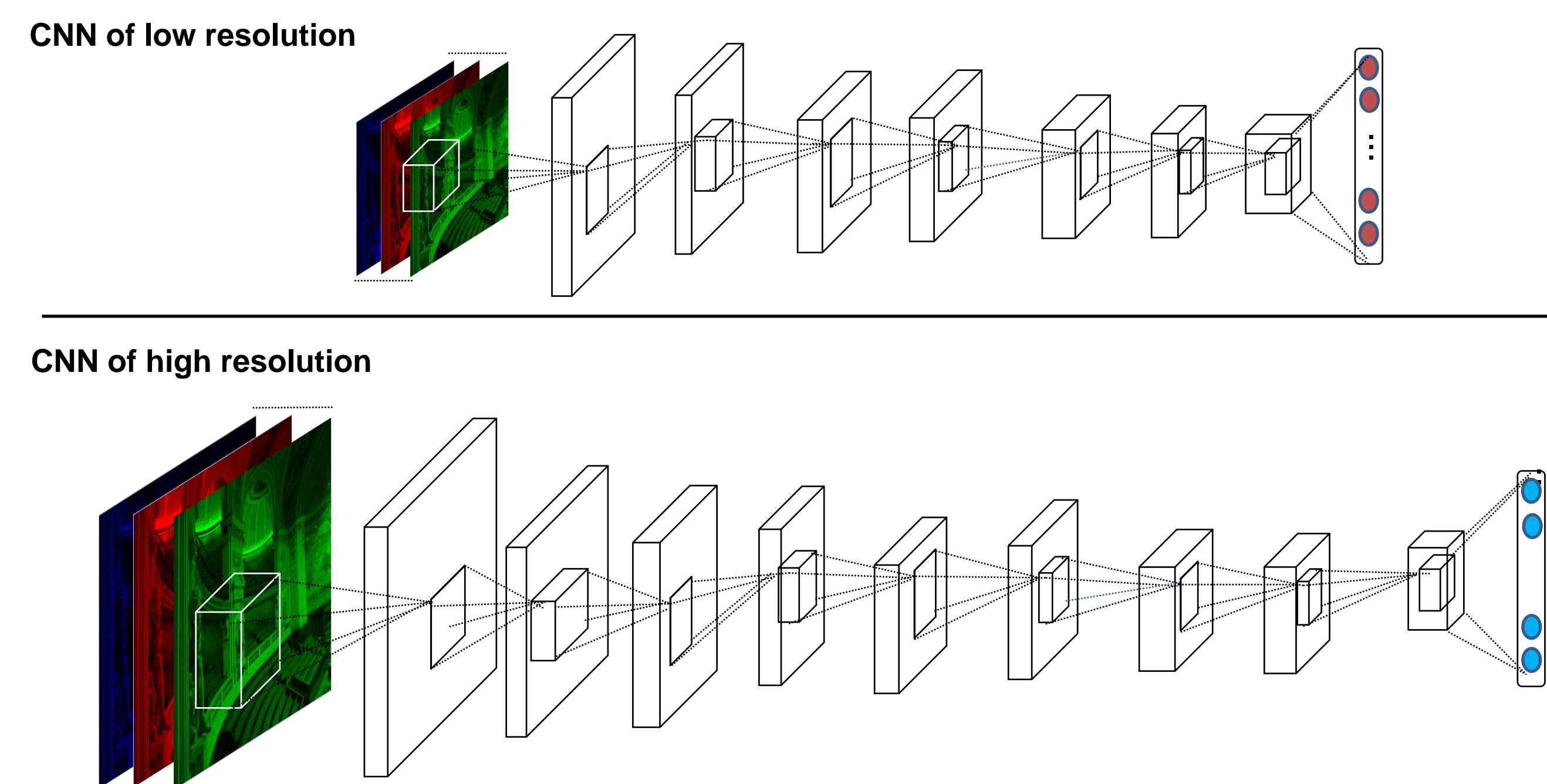
Highlights

- Large scale scene datasets with many categories come along with **increased ambiguity** between the class labels (e.g. baseball field vs. stadium baseball).
- **Knowledge guided disambiguation** aims to regularize CNN training with extra knowledge and improve the generalization capacity.
- **Multi-Resolution CNNs** take images of different sizes as input and capture visual information from different levels.

Label Ambiguity



Multi-Resolution CNNs



Challenge Solutions and Results

- **Architectures:**
Low resolution: image (256*256), crop(224*224), inception2 network [2]
High resolution: image (384*384), crop(336*336), inception2+2 convs
- **Knowledge networks:**
Object nets: inception2 trained with ImageNet
Scene nets: inception2 trained with Places205
 Currently, knowledge disambiguation only for low resolution CNNs
- **Implementation details:**
 Resample images to balance the class distribution
 Data augmentation: fixed crop, scale jittering, horizontal flipping [1,6]
Toolbox: we use a multi-GPU extension of Caffe, which is publicly available: <https://github.com/yjxiong/caffe.git> [6]

Results on Places2 validation dataset

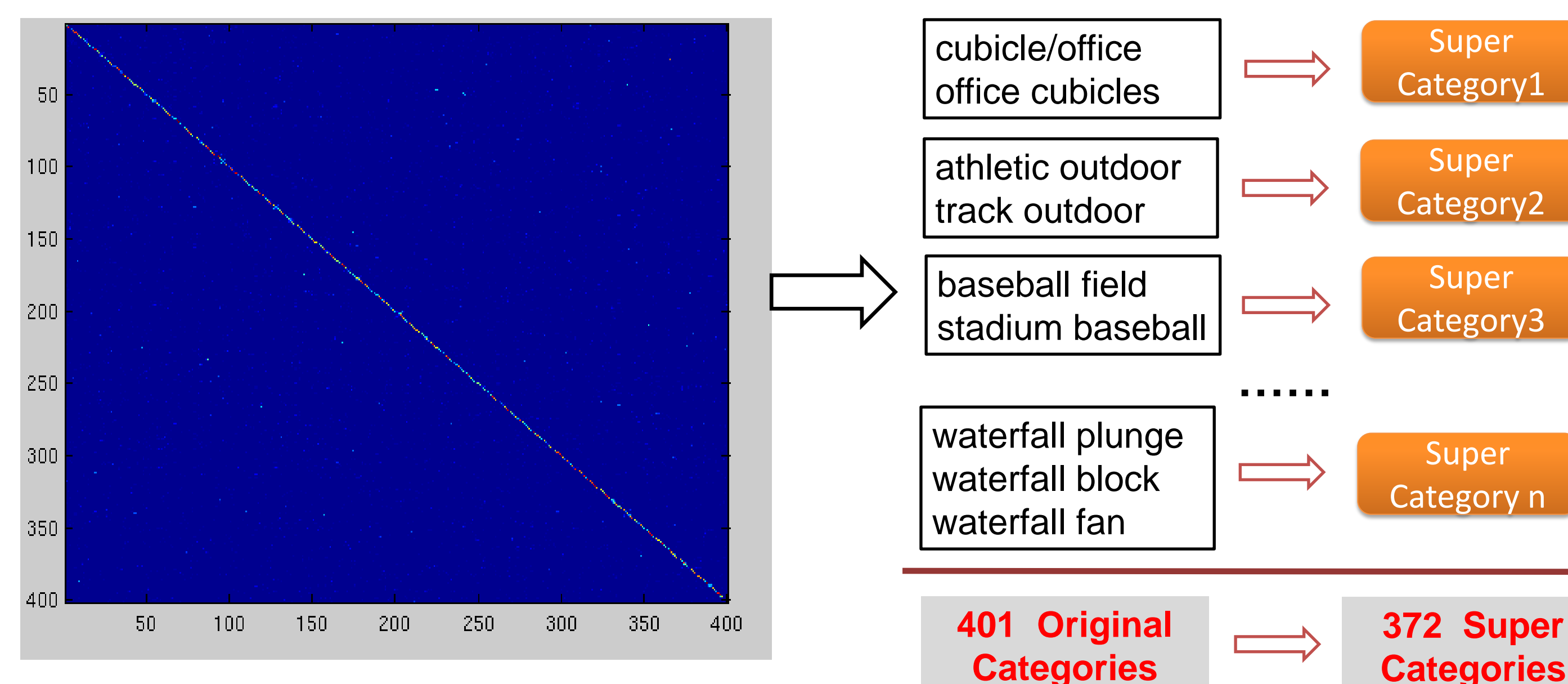
Architecture	Training Method	Top5
Inception2 (256*256) [2]	Normal training	17.4
Inception2 (256*256) [2]	Normal training (1024 batch size)	17.2
Inception2 (256*256) [2]	Fine tune (Places 205)	17.9
Inception2 (256*256) [2]	Locally supervised	17.2
Inception2 (256*256) [2]	Confusion matrix knowledge	17.3
Inception2 (256*256) [2]	Scene net knowledge	16.7
Inception2 (256*256) [2]	Object net knowledge	17.4
Inception2+2convs (384*384)	Normal training	16.6
VGGNet-16 (256*256) [3]	Normal training	17.5
VGGNet-19 (256*256) [3]	Normal training	17.8
MSRA B Model (256*256) [4]	Normal training	17.9

Results on Places2 test dataset

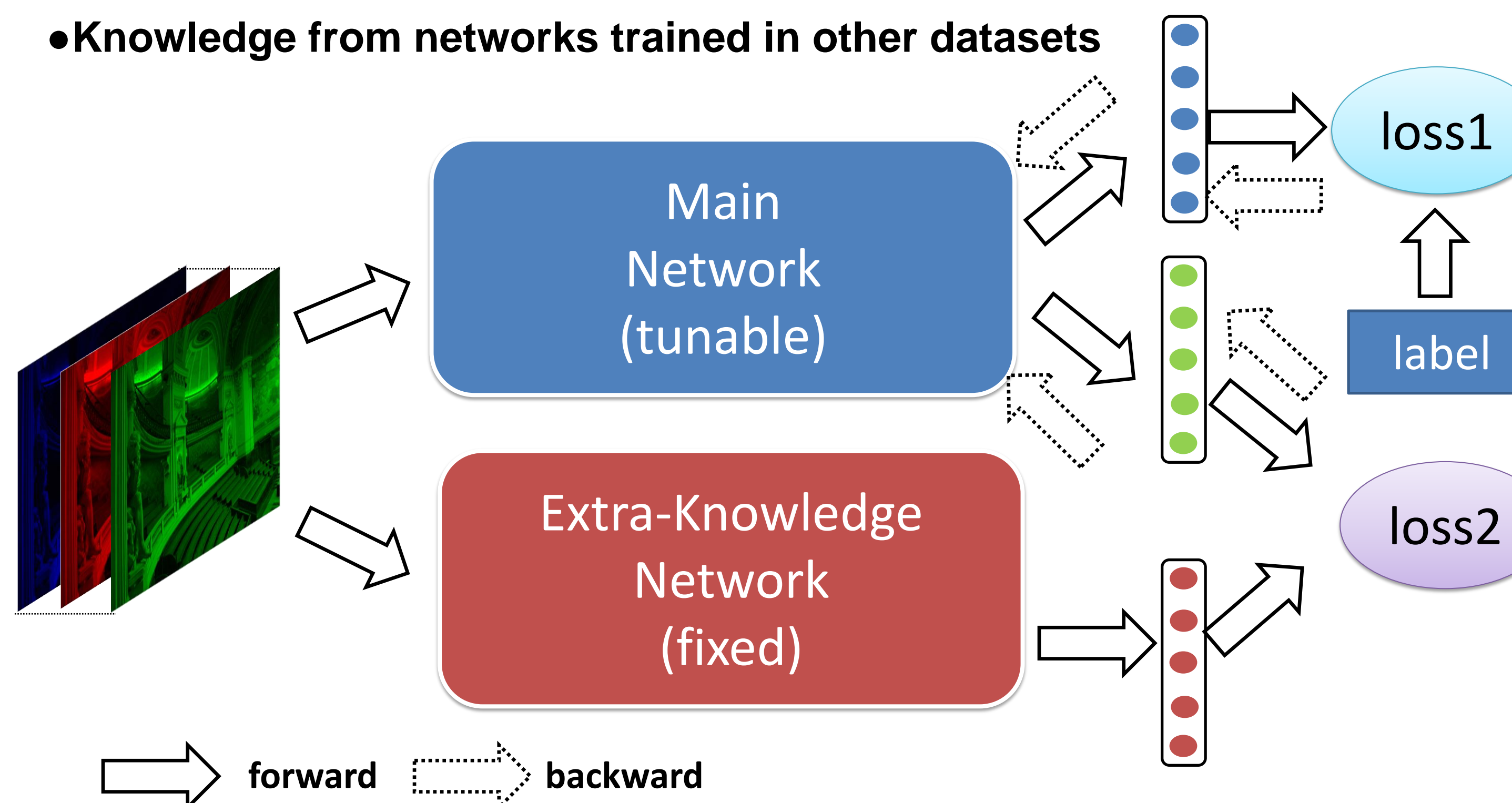
Rank	Team	Top5	Rank	Team	Top5
1	WM	16.9	5	NTU_Rose	19.3
2	Our (2 model)	17.6	6	Mitsubishi Electric	19.4
	Our (best)	17.4	7	HiVision	20.0
3	Qualcomm	17.6	8	DeepSEU	20.0
4	Trimps-Soushen	18.0	9	Gator Vision	20.3

Knowledge Guided Disambiguation

Knowledge from confusion matrix on validation dataset



Knowledge from networks trained in other datasets



References

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.
- [2] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR, abs/1502.03167, 2015.
- [3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.
- [4] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. CoRR, abs/1502.01852, 2015.
- [5] Hinton, Geoffrey, Vinyals, Oriol, and Dean, Jeff. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
- [6] L. Wang, Y. Xiong, Z. Wang, and Y. Qiao. Towards good practices for very deep two-stream convnets. CoRR, abs/1507.02159, 2015.