











- Bischof, H., Schneider, W., and Pinz, A., 1992. Multispectral classification of landsat images using neural networks. *IEEE Trans. on Geoscience and Remote Sensing*, 30(3):482–490.
- Castelluccio, M., Poggi, G., Sansone, C., and Verdoliva, L. 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv preprint arXiv:1508.00092.
- Farabet, C., Couprie, C., Najman, L and LeCun. Y., 2013. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Analysis. Machine. Intelligence*, 35(8), pp.1915–1929.
- Friedl, M. A and Carla, E. B., 1997. Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3), pp. 399-409.
- Gerke, M., ISPRS 2D Semantic Labeling Contest. <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html>. 1, 5 (20 August. 2015)
- Gerke, M., 2015. Use of the stair vision library within the ISPRS 2D semantic labelling benchmark (Vaihingen). Technical report, University of Twente, 2015.
- Girshick, R., Donahue, J., Darrell, T., and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. pp. 580-587.
- Khaleghi, B., Khamis, A., Karray, F. O., Razavi, S. N., 2013. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1), pp.28-44.
- Krizhevsky, K., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in *conference on advances in neural information processing systems*, pp.1097-1105.
- Liu, K., Skibbe, H., Schmidt, T., Blein, Z., Palme, K., Brox, T., Ronneberger, O., 2014. Rotation-invariant hog descriptors using fourier analysis in polar and spherical coordinates. *International Journal of Computer Vision*, 106(3), pp.342–364.
- Niemeyer, J., Wegner, J., Mallet, C., Rottensteiner, F and Soergel, U., 2011. Conditional random fields for urban scene classification with full waveform lidar data. In *Photogrammetric Image Analysis*, volume, 6952, pp. 233–244.
- Paisitkriangkra, S., Sherrahy, J., Janneyz, P and Van-Den Hengel, A., 2014. Report on the ISPRS 2D Semantic Labeling Contest. “Effective Semantic Pixel labelling with Convolutional Networks and Conditional Random Fields”, [http://www.itc.nl/external/ISPRS\\_WGIII4/ISPRSIII\\_4\\_Test\\_results/papers/DSTO\\_.pdf](http://www.itc.nl/external/ISPRS_WGIII4/ISPRSIII_4_Test_results/papers/DSTO_.pdf) (28 May. 2015).
- Saeidi, V., Pradhan, B., Idrees, M. O., and Abd Latif, Z., 2014. Fusion of airborne LiDAR with multispectral spot 5 image for enhancement of feature extraction using Dempster–Shafer theory. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10), pp.6017-6025.
- Razavian, A.S., Azizpour, H., Sullivan, J and Carlsson, S., 2014. CNN features off the-shelf: an astounding baseline for recognition. In Proc. IEEE Conf. Computer. Vision. Pattern Recognition, Columbus, OH, pp.512-519.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information fusion*, 6(4), pp.283-300.
- Tabib Mahmoudi, F., Samadzadegan, F., Reinartz, P., 2015. Object recognition based on the context aware decision-level fusion in multiviews imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(1), pp.12-22.
- Wei, Y., Yao, W., Wu, J., Schmitt, M., Stilla, U. 2012. Adaboost-based feature relevance assessment in fusing lidar and image data for classification of trees and vehicles in urban scenes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(7), pp.323-328.
- Yao, W., Hinz, S., Stilla, U., 2011. Extraction and motion estimation of vehicles in single-pass airborne LiDAR data towards urban traffic analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), pp.260-271.

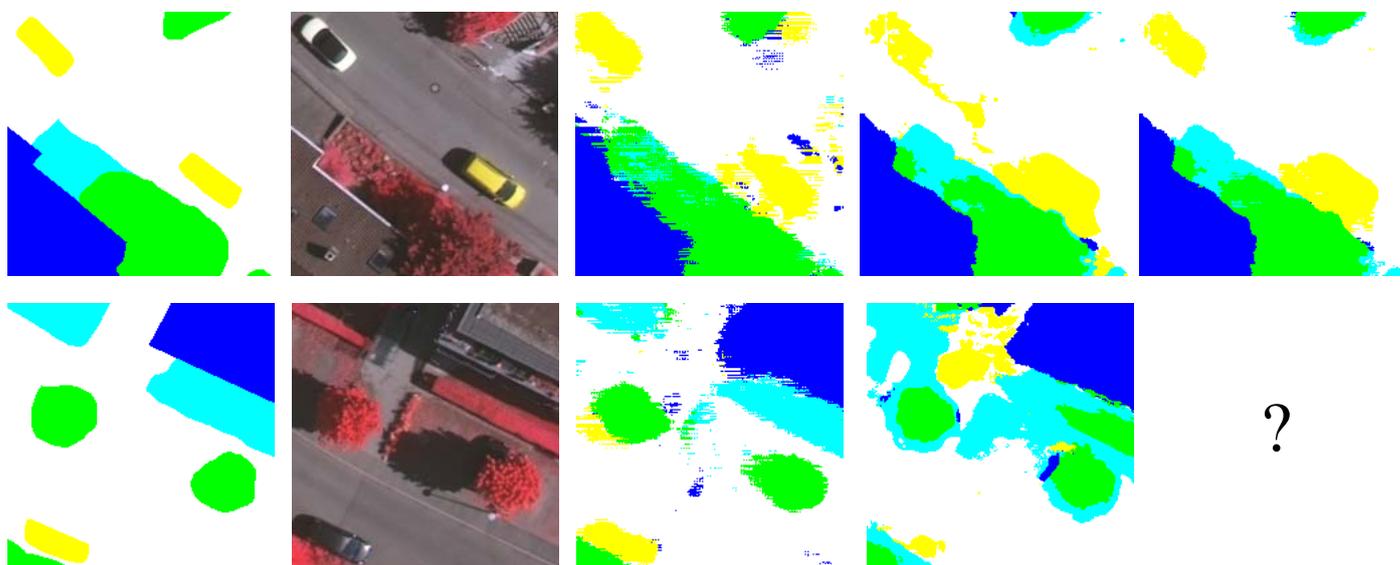


Figure 3. Two examples of classification results for complementary (top) and conflicting (bottom) scenes. From left to right: ground truth, input image CNN labelling, combined classifier labelling