This checklist describes the algorithm we used for our submission to the CATARACTS challenge.

Checklist

- Contact details: evangello.flouty@touchsurgery.com
- Name of the algorithm: CatResNet
- Give the overall structure of the algorithm and briefly describe each step in this structure. For the task of tool presence detection we used ResNet-152 [1] towards multi-label classification. We initialized the network using the weights pretrained on the ImageNet dataset [2], except for the output fully connected layer of 21 output nodes (one for each tool class), which is specific to the task at hand and which was initialized with a gaussian distribution of mean 0 and standard deviation 0.01. The network was fine-tuned using the CATARACTS data. To train our model we extracted frames from the 25 train videos with 3fps and discarded half of the frames that do not contain any tool. In training time we re-shaped the input frames to 398 × 224, performing random horizontal flips and rotations with mirror padding to avoid over-fitting. We then do a bottom center crop to the image preserving one of the dimensions to reduce distortions as much as possible. We used SGD with a mini-batch of 16, learning rate of 10^{-4} and momentum of 0.9 for a total of 10,000 iterations.
- Does your system use additional training data? If so, describe the characteristics of that training data. We only used the CATARACTS dataset.
- Do predictions for a given frame solely rely on that frame? Or do they also rely on previous frames? on following frames? Predictions for each frame rely only on the current frame.
- Give computation times (number of frames processed per second). We processed 15 frames per second.
- If the algorithm has been tested on other datasets, you could consider including those results. We haven't tested our algorithm on other datasets.
- References
    - [1] Kaiming, H., Xiangyu Z., Shaoqing R., et al.: Deep Residual Learning for Image Recognition. CVPR, 770–778 (2016)
    - [2] Russakovsky, O., Deng, J., Su, H., et al: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3), 211–252 (2015) 1