M2CAI Workflow Challenge: Convolutional Neural Networks with Time Smoothing and Hidden Markov Model for Video Frames Classification

Remi Cadene, Thomas Robert, Nicolas Thome, Matthieu Cord

Sorbonne Universités, UPMC Univ Paris 06, LIP6, Paris, France

Context

Goal: Surgical video frames classification

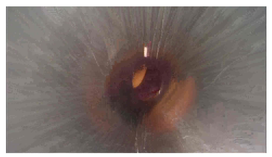
▷ Videos of size 1920x1080 Shot at 25 frames per second at IRCAD research center in Strasbourg, France

▷ 27 training videos

▷ 15 testing videos

▷ 8 classes

Clean image



Noisy image

▷ Online prediction: $P(y|x_i, x_{i-1}, x_{i-2}, ...)$

▷ Usefull to

- ▷ Monitor surgeons
- ▷ Trigger automatic actions

Deep Learning Methods

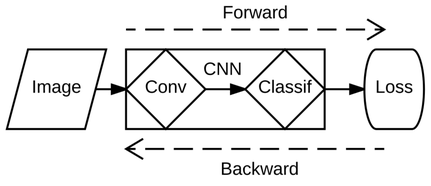
Random split + sampling (1f/s):

▷ Training set: 22 videos (59,493 images)

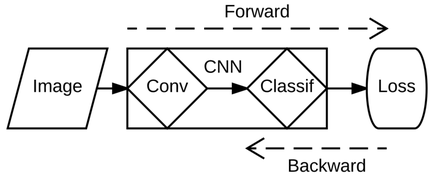
▷ Validation set: 5 videos (8,062 images)

▷ Testing set: 15 videos (28,732 images)

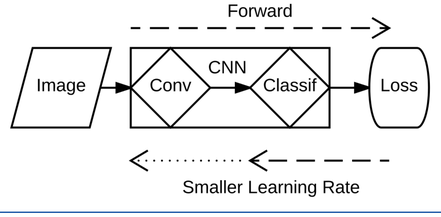
"From Scratch" : End-to-End Learning



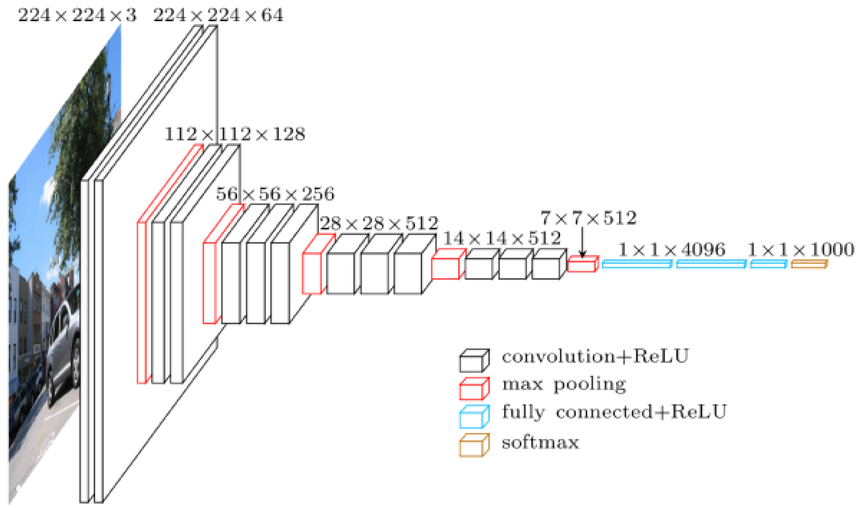
"Extraction" : Pre-trained CNN



"Fine-Tuning" : Both approaches !



Deep Learning architecture and scores



Model	Type	Accuracy (%)
InceptionV3	Extraction (repres. of ImageNet)	60.53
InceptionV3	From Scratch (repres. of M2CAI)	69.13
InceptionV3	Fine-tuning (both representations)	79.06
ResNet200	**Fine-tuning (both representations)**	**79.24**

Experiments

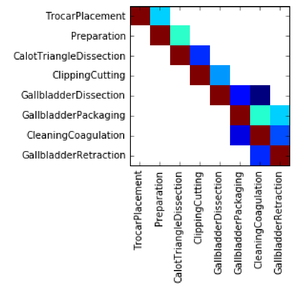
Temporal Method	Accuracy Val (%)	Jaccard Val	Jaccard Test
No Smoothing	79.24	–	–
Avg Smoothing	85.97	74.67	–
Avg + HMM Online	**88.90**	**81.60**	**71.9**
Avg + HMM Offline	93.47	87.59	–

Hidden Markov Model

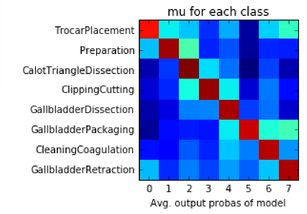
▷ Initial state probabilities

▷ Matrix of probabilities of transition between states

▷ Gaussian parameters for emissions of observations (mean and co-variance matrix)



(b)



(c mean)

Conclusion

Results are reproductible:

`github.com/Cadene/torchnet-m2caiworkflow`

[1] Oquab et al. Is object localization for free? *CVPR*, 2015.
[2] Durand et al. MANTRA. *ICCV*, 2015.
[3] Parizi et al. Automatic discovery of parts. *ICLR*, 2015.
[4] Gong et al. Multi-scale orderless pooling. *ECCV*, 2014.