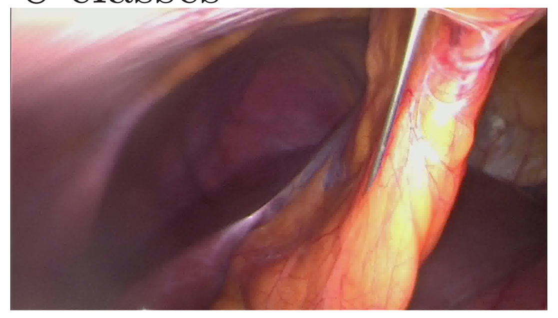
M2CAI Workflow Challenge: Convolutional Neural Networks with Time Smoothing and Hidden Markov Model for Video Frames Classification

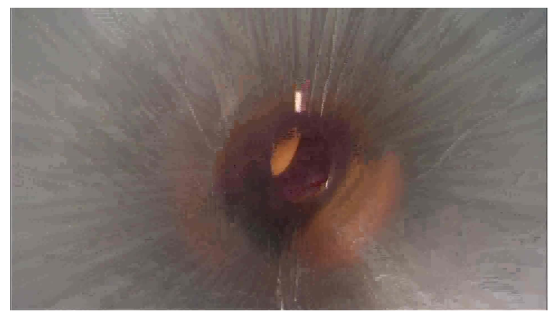
Remi Cadene, Thomas Robert, Nicolas Thome, Matthieu Cord
Sorbonne Universités, UPMC Univ Paris 06, LIP6, Paris, France

Context

Goal: Surgical video frames classification
- ▷ Videos of size 1920x1080 Shot at 25 frames per second and at IRCAD research center in Strasbourg, France
- ▷ 27 training videos
- ▷ 15 testing videos
- ▷ 8 classes



Clean image



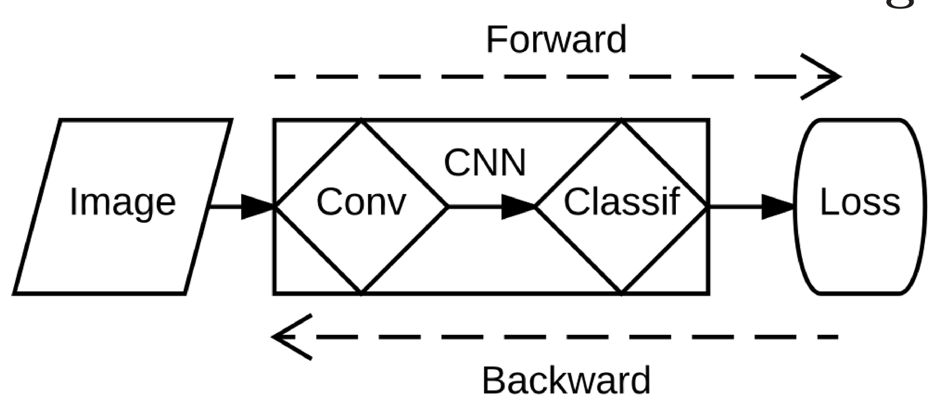
Noisy image

- ▷ Online prediction: $P(y|x_i, x_{i-1}, x_{i-2}, ...)$
- ▷ Usefull to
  - ▷ Monitor surgeons
  - ▷ Trigger automatic actions

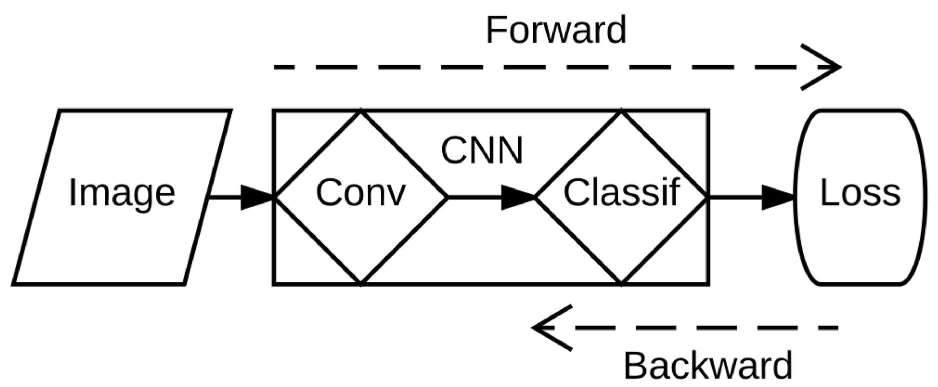
Deep Learning Methods

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)

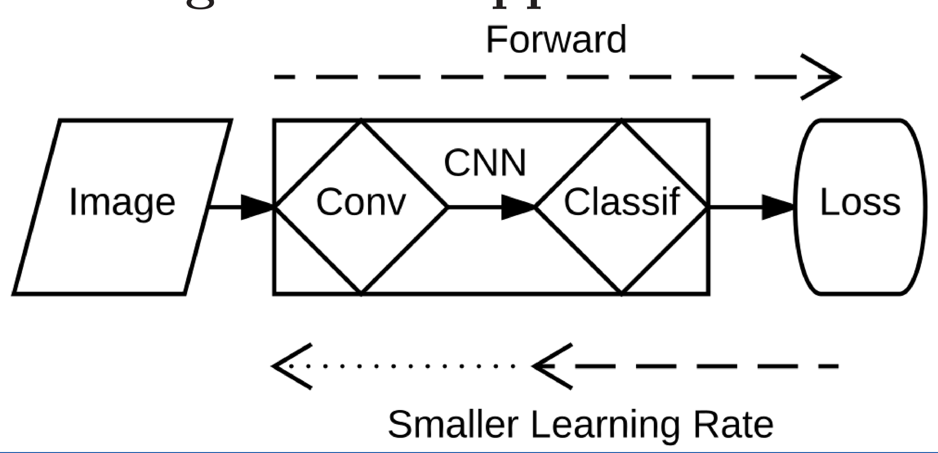
"From Scratch" : End-to-End Learning



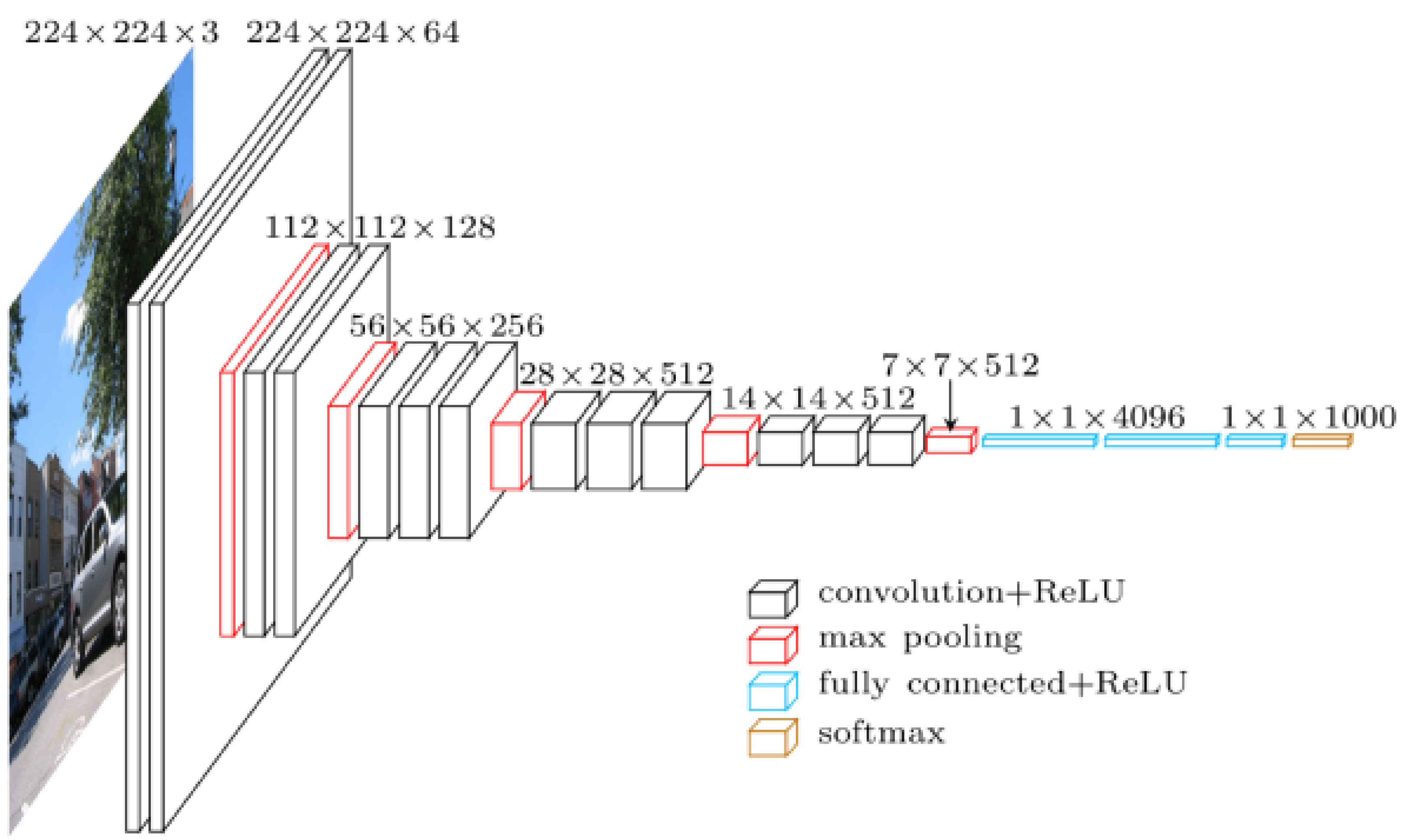
"Extraction" : Pre-trained CNN

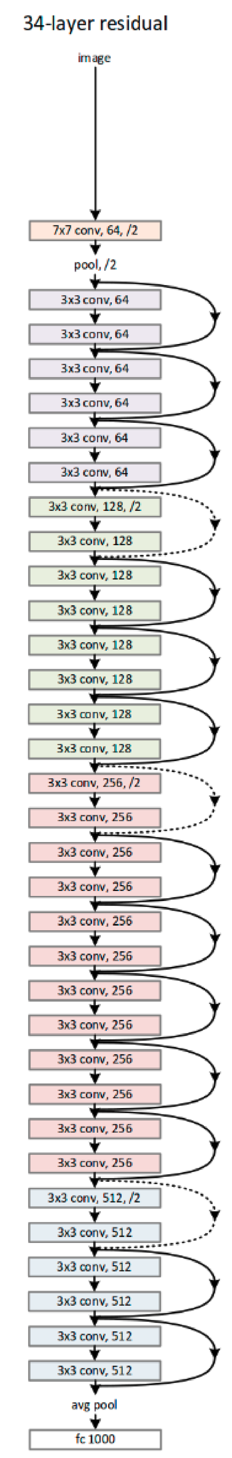


"Fine-Tuning" : Both approaches !



Comparison of Deep Learning architectures

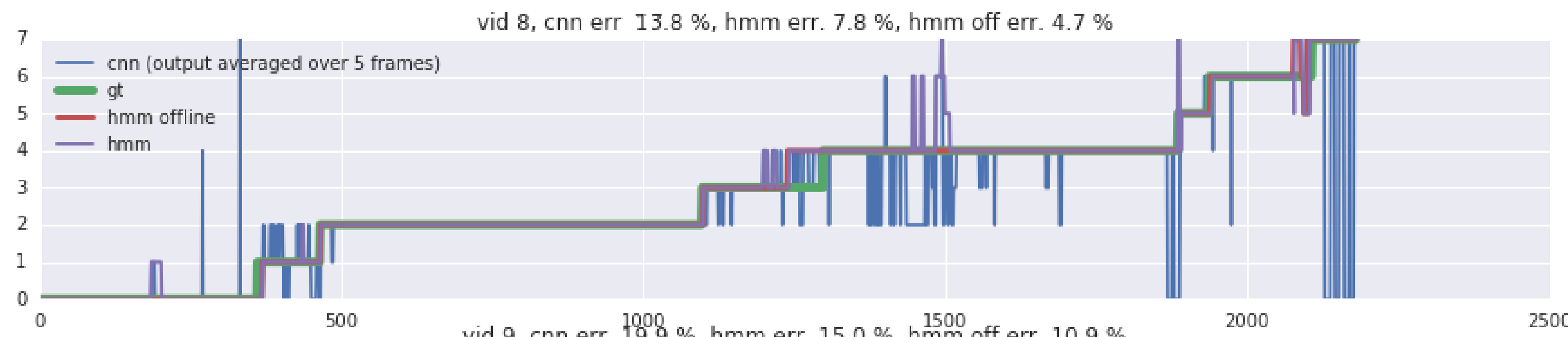


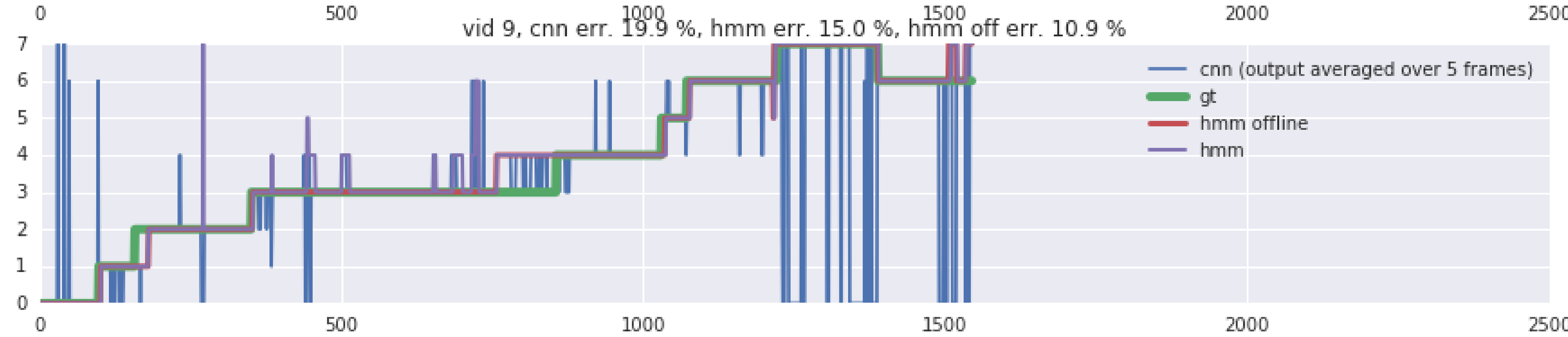


Model	Type	Accuracy (%)
InceptionV3 [4]	Extraction (repres. of ImageNet)	60.53
InceptionV3 [4]	From Scratch (repres. of M2CAI)	69.13
InceptionV3 [4]	Weldon [3] (both representations)	78.18
InceptionV3 [4]	Fine-tuning (both representations)	79.06
ResNet200 [2]	**Fine-tuning (both representations)**	**79.24**

Comparison of Smoothing methods

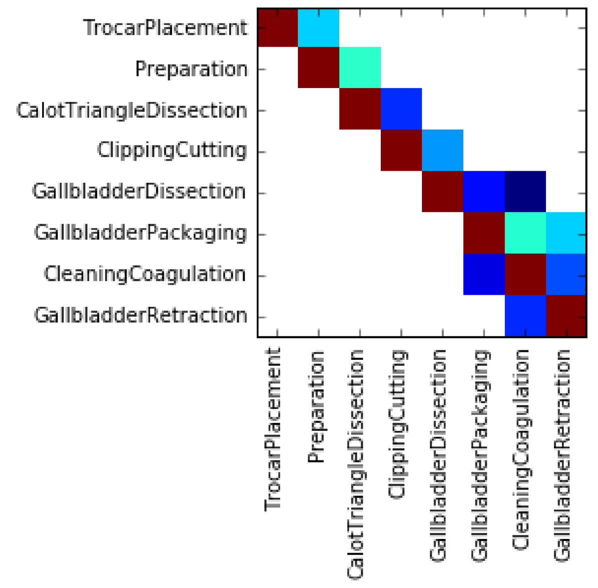
Temporal Method	Accuracy Val (%)	Jaccard Val	Jaccard Test
No Smoothing	79.24	–	–
Avg Smoothing	85.97	74.67	–
Avg + HMM Online	**88.90**	**81.60**	**71.9**
Avg + HMM Offline	93.47	87.59	–

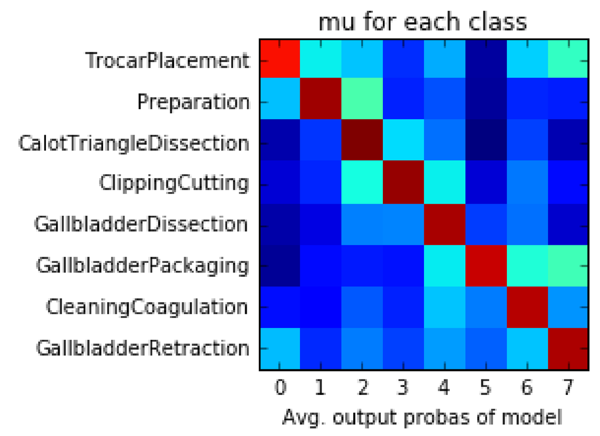




Hidden Markov Model

- ▷ Initial state probabilities
- ▷ Matrix of probabilities of transition between states
- ▷ Gaussian parameters for emissions of observations (mean and co-variance matrix)





Conclusion

Future works:
- ▷ Fine tuning CNN on full trainset (not only 80%)
- ▷ Ensembling several fine tuned CNNs

Results are reproducible with Torch7:

`github.com/Cadene/torchnet-m2caiworkflow`
`arxiv.org/abs/1610.05541`

Code

Paper

1 Cadene et al. Master's Thesis : Deep Learning for Visual Recognition. *arXiv*, 2016.

2 He K. et al. Identity Mappings in Deep Residual Networks. *ECCV*, 2016.

3 Durand et al. WELDON: Weakly Supervised Learning of Deep Convolutional Neural Networks. *CVPR*, 2016.

4 Szeggedy et al. Rethinking the inception architecture for computer vision *CVPR*, 2015.

5 Diederik et al. Adam: A Method for Stochastic Optimization. *ICLR*, 2014.

6 Simonyan et al. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2014.