The Contributions of Deep Learning to Computer Vision: Appplication to Medical Images

Nicolas Thome
Professor at Conservatoire National des Arts et Métiers (Cnam)
Machine Learning Team (MSDMA) - Computer Science Department
Prenom.Nom@cnam.fr - http://cedric.cnam.fr/~thomen/

April 4, 2018

Context: Big Data

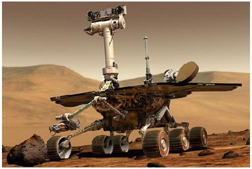
- Superabundance of visual data: images, videos, *etc*

BBC: 2.4M videos

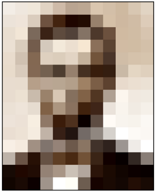
Social media,
e.g. Facebook: 1B each day

100M monitoring cameras

- Obvious need for **Visual Recognition**
- Huge number of applications: mobile visual search, medical imaging, robotics, autonomous driving, augmented reality, *etc*



Challenge: filling the semantic gap



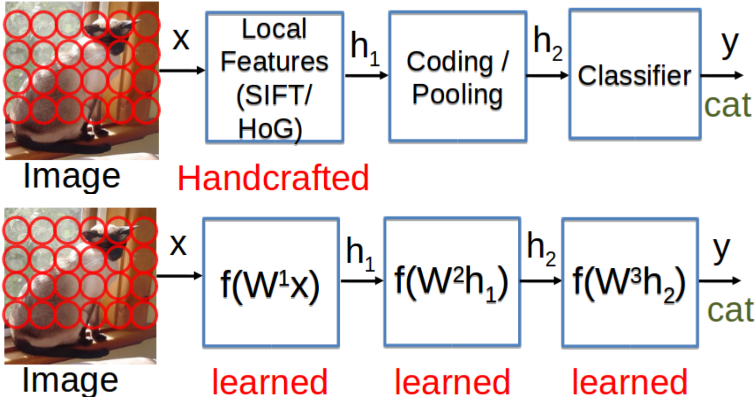
What we perceive *vs*
What a computer sees

243	239	240	225	206	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	108	208	208	215
235	217	115	212	243	236	247	139	91	209	208	211
233	208	131	222	219	226	196	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
232	232	182	186	184	179	159	123	93	232	235	235
232	236	201	154	216	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	59	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	149	109	138	65	47	156	239	255
190	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

- Illumination variations
- View-point variations
- Deformable objects
- intra-class variance
- *etc*

⇒ **How to design "good" intermediate representations ?**

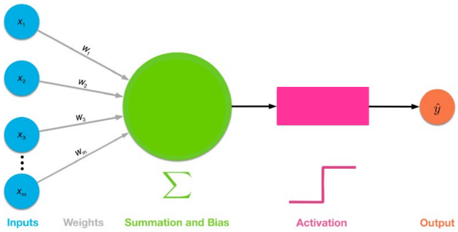
- Before DL:
  handcrafted intermediate representations for each domain
- Since DL:
  Representation Learning



Outline

1. Convolutionnal Neural Networks (ConvNets)
2. Deep learning for Medical Images

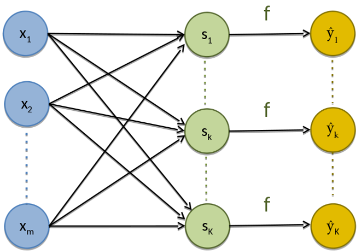
- The formal neuron



x_i: inputs
w_i, b: weights
f: activation function
y: output of the neuron

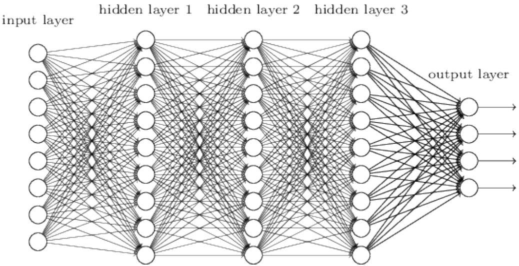
$$y = f(w^\top x + b)$$

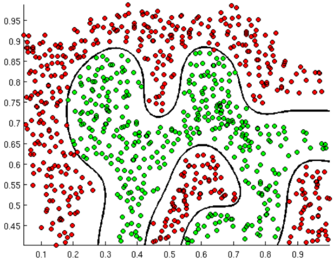
- Stacking several formal neurons ⇒ **Perceptron**



The Multi-Layer Perceptron (MLP)

- Perceptron: limited to linear decision boundaries
- Stacking layers of neural networks ⇒ more complex and rich functions

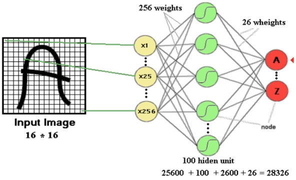


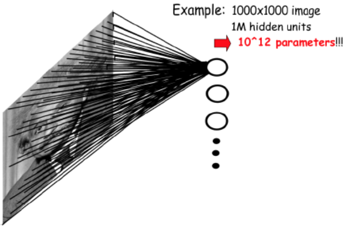


- Basis of the "deep learning" field
- **All parameters trained with backpropagation with class labels**

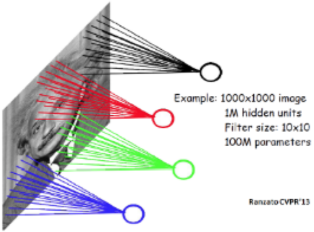
Convolutional Neural Networks (ConvNets)

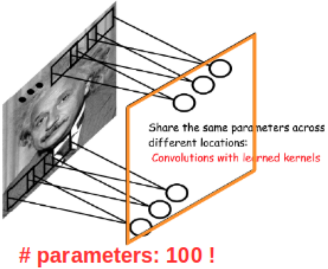
- Scalability issue with Fully Connected Networks (MLP) + no local information!





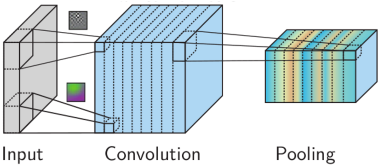
- **<u>ConvNets:</u> sparse connections, shared weights = compact + local features**



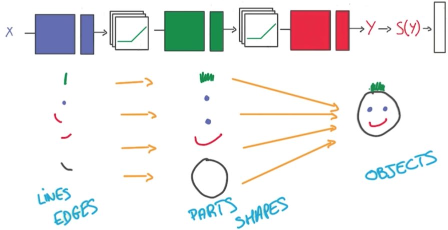


Convolutional Neural Networks (ConvNets)

- Convolution on tensors, *i.e.* multidimensional arrays: T of size $W \times H \times D$
  - Convolution: $C[T] = T'$, T' tensor of size $W' \times H' \times K$
  - Each filter locally connected with shared weights (K number of filters)
- **Elementary block: Convolution + Non linearity (*e.g.* ReLU)+ pooling**

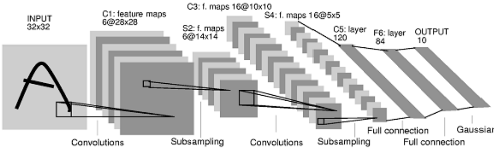


- **Stacking several Blocks:** intuitive hierarchical information extraction

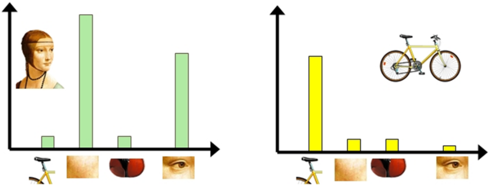


Deep Learning History

- 80's: training Convolutionnal Neural Networks (CNN) with back-propagation ⇒ postal code reading [LeCun et al., 1989]

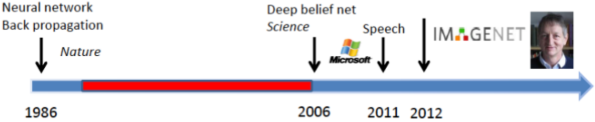


- 90's: golden age of kernel methods, NN = black box
- 2000's: BoW + SVM : state-of-the-art CV



Deep Learning History

- Deep learning revival in 2012: outstanding success of ConvNets in ImageNet [Krizhevsky et al., 2012]



Rank	Name	Error rate	Description
1	**U. Toronto**	0.15315	Deep learning
2	U. Tokyo	0.26172	Hand-crafted features and learning models. Bottleneck.
3	U. Oxford	0.26979	
4	Xerox/INRIA	0.27058	

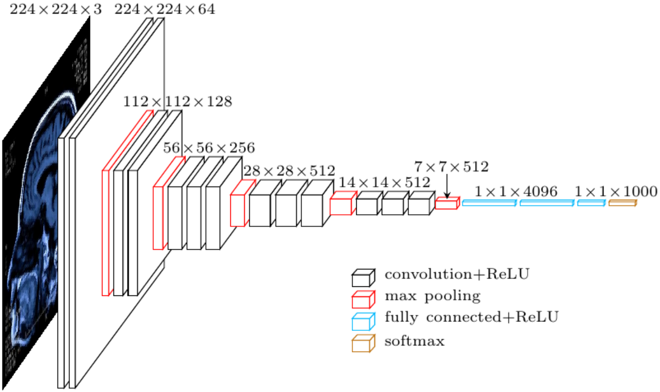
- **Two main practical reasons:**
  1. Huge number of labeled images (10^6 images)
  2. GPU implementation for training

IMAGENET

Outline

Deep Learning (DL) for Medical Image Diagnostic

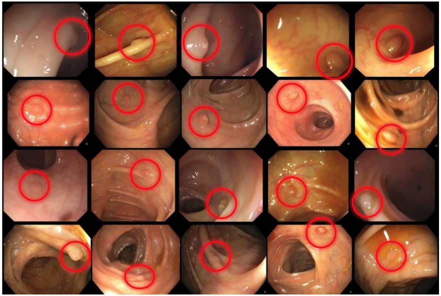
- Deep ConvNets require large-scale annotated datasets
- **BUT:** Transferring Representations learned from ImageNet
  Extract layer (fixed-size vector) ⇒ **"Deep Features" (DF)**

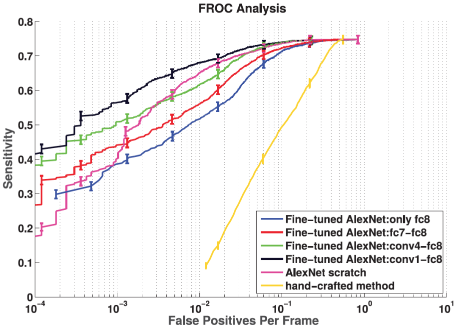


- **Now state-of-the-art for any visual recognition task** [Azizpour et al., 2016]
- **DF very robust to domain shifts, *e.g.* medical images**

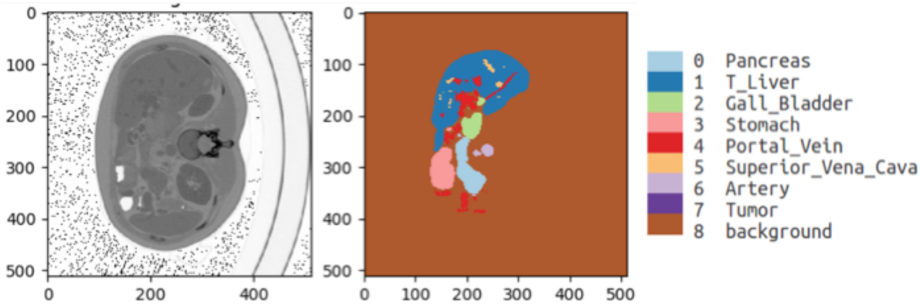
Deep Learning (DL) for Medical Image Diagnostic

- DL & ConvNets: performance boost for classification in medical images
- Transfer & fine-tuning (ImageNet), *e.g.* Polyp Detection [Tajbakhsh et al., 2016]
- ConvNets trained from scratch, *e.g.* Mammography Classification [Kooi et al., 2017]
- ConvNets: winners of recent challenges based on deep learning: Mammography, Melanoma Detection, *etc*



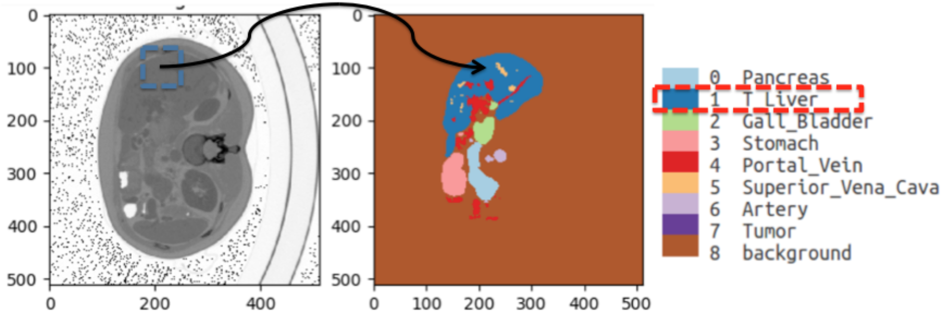


- **Semantic segmentation:** assigning a label to each image pixel

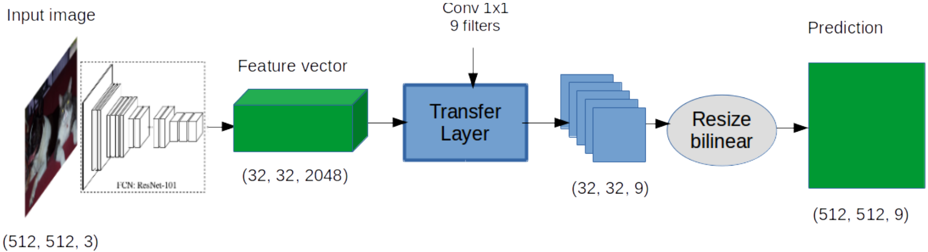


Deep Learning for Medical Image Semantic Segmentation

- **Deep Learning segmentation:** classifying image regions around each pixel



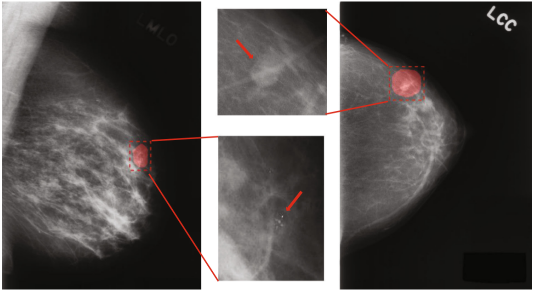
- Standard computer vision models models based on Fully Convolutional Networks (FCN)



- FCN base models for many state-of-the-art methods segmentation methods, *e.g.* leading approach in Liver Tumor Segmentation (LiTS'17) challenge [Li et al., 2017]

Deep Learning for Medical Images

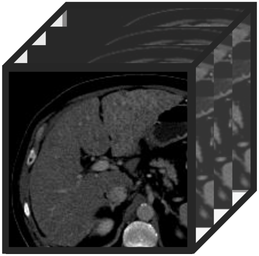
- Successful exportation of DL solutions boost performances... **BUT**
- **... Medical images very different from natural images:**
  - **Discriminative pattern often tiny**, *e.g.* Mammography 0.5% – 1.2% cancer pixel [Akselrod-Ballin et al., 2017] *vs* $>$ 50% ImageNet or $>$ 30% VOC
  - $\Rightarrow$ **Strong imbalance** between $\oplus$ and $\ominus$ (background) classes



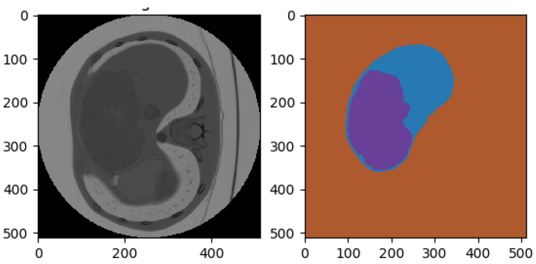
Calcification(0.5%) Mass(1.2%)

Deep Learning for Medical Images

- Successful exportation of DL solutions boost performances**... BUT**
- **... Medical images very different from natural images:**
  - **3D volumes** *vs* 2D Images
  - Hierarchical / nested detection or organs, *e.g.* tumor inside liver

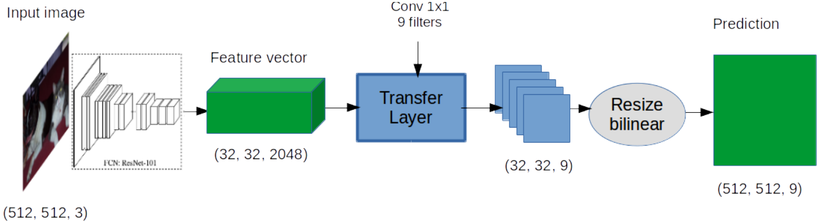


Input

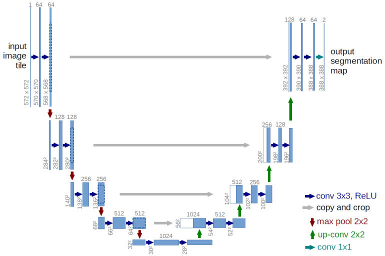


Specific Deep Learning Architectures for Medical Images

Resolution loss through the network

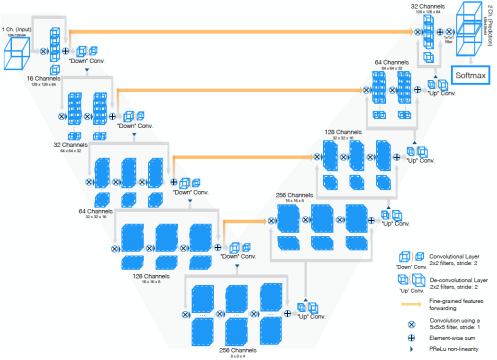


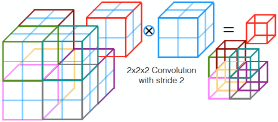
- **Introduction of skip connections** in U-Net [Ronneberger et al., 2015]



Representation Learning with 3D Inputs?

- Use 3D convolution, *e.g.* V-Net [Milletari et al., 2016], 3D U-Net [Çiçek et al., 2016] or [Lu et al., 2017]

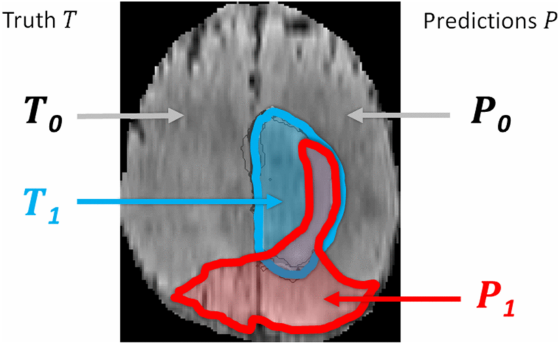




Class imbalance

Use a specific loss function, *e.g.*

- Weighted cross entropy, U-Net [Ronneberger et al., 2015]
- Dice score, V-Net [Milletari et al., 2016] or [Fidon et al., 2017, Sudre et al., 2017]

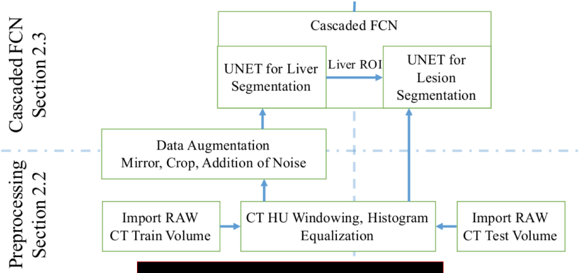


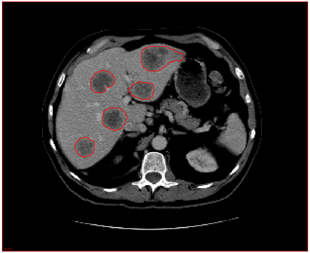
$$\mathbf{s} = \frac{2|T \cap P|}{|T| + |P|}$$

Specific Training Schemes for Medical Images

Exploit *prior* knowledge between organs, *e.g.* tumors only in liver

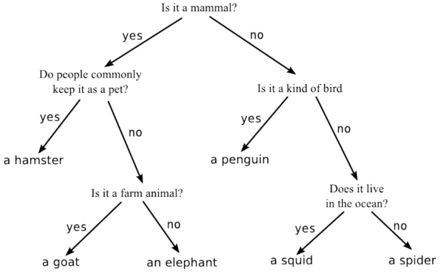
- Cascaded FCNNs for liver-tumor segmentation [Christ et al., 2016]

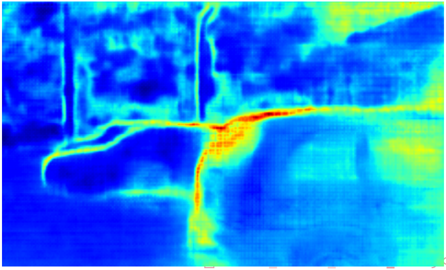




Conclusion

- Deep Learning & ConvNets: state-of-the-art solutions for medical image analysis
  - **Representation learning** ⇒ better visual features
- Exporting solutions from computer vision: transfer for classification, RPN for localization, FCN for segmentation, *etc*
  - Some adaptation required: spatial resolution, class imbalance, 3D data, *etc*
- Other crucial steps for deploying DL solution in Healthcare: uncertainty estimate and explainability ⇒ vanilla DL models poor at these tasks
  - Some preliminary solutions for uncertainty [Gal and Ghahramani, 2016] and explainability [Frosst and Hinton, 2017]





Thank you for your attention!

Questions?

References I

[Akselrod-Ballin et al., 2017] Akselrod-Ballin, A., Karlinsky, L., Hazan, A., Bakalo, R., Horesh, A. B., Shoshan, Y., and Barkan, E. (2017).
Deep learning for automatic detection of abnormal findings in breast mammography.
In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 321–329. Springer.

[Azizpour et al., 2016] Azizpour, H., Razavian, A. S., Sullivan, J., Maki, A., and Carlsson, S. (2016).
Factors of transferability for a generic convnet representation.
IEEE Trans. Pattern Anal. Mach. Intell., 38(9):1790–1802.

[Christ et al., 2016] Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M., Sommer, W. H., Ahmadi, S., and Menze, B. H. (2016).
Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields.
In *MICCAI (2)*, volume 9901 of *Lecture Notes in Computer Science*, pages 415–423.

[Çiçek et al., 2016] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016).
3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.
In *MICCAI (2)*, volume 9901 of *Lecture Notes in Computer Science*, pages 424–432.

[Fidon et al., 2017] Fidon, L., Li, W., García-Peraza-Herrera, L. C., Ekanayake, J., Kitchen, N., Ourselin, S., and Vercauteren, T. (2017).
Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks.
In *BrainLes@MICCAI*, volume 10670 of *Lecture Notes in Computer Science*, pages 64–76. Springer.

[Frosst and Hinton, 2017] Frosst, N. and Hinton, G. E. (2017).
Distilling a neural network into a soft decision tree.
In *CEx@AI*IA*, volume 2071 of *CEUR Workshop Proceedings*. CEUR-WS.org.

[Gal and Ghahramani, 2016] Gal, Y. and Ghahramani, Z. (2016).
Dropout as a Bayesian approximation: Representing model uncertainty in deep learning.
In *Proceedings of the 33rd International Conference on Machine Learning (ICML-16)*.

References II

[Kooi et al., 2017] Kooi, T., Litjens, G. J. S., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., and Karssemeijer, N. (2017).
Large scale deep learning for computer aided detection of mammographic lesions.
Medical Image Analysis, 35:303–312.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).
Imagenet classification with deep convolutional neural networks.
In *Advances in neural information processing systems*, pages 1097–1105.

[LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989).
Backpropagation applied to handwritten zip code recognition.
Neural computation, 1(4):541–551.

[Li et al., 2017] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C., and Heng, P. (2017).
H-denseunet: Hybrid densely connected unet for liver and liver tumor segmentation from CT volumes.
CoRR, abs/1709.07330.

[Lu et al., 2017] Lu, F., Wu, F., Hu, P., Peng, Z., and Kong, D. (2017).
Automatic 3d liver location and segmentation via convolutional neural network and graph cut.
Int. J. Computer Assisted Radiology and Surgery, 12(2):171–182.

[Milletari et al., 2016] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016).
V-net: Fully convolutional neural networks for volumetric medical image segmentation.
In *IEEE International Conference on 3DVision (arXiv:1606.04797)*.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015).
U-Net: Convolutional Networks for Biomedical Image Segmentation.
In *MICCAI (3)*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer.

[Sudre et al., 2017] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Cardoso, M. J. (2017).
Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations.
In *DLMIA/ML-CDS@MICCAI*.

References III

[Tajbakhsh et al., 2016] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016).
Convolutional Neural Networks for Medical Image Analysis: Fine Tuning or Full Training?
IEEE Transactions on Medical Imaging, PP(99):1–1.