BASAL CELL CARCINOMA DETECTION IN FULL FIELD OCT IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Diana Mandache[1], Eugénie Dalimier[2], John R Durkin[3], Claude Boccara[2], Jean-Christophe Olivo-Marin[1], Vannary Meas-Yedid[1]

[1] Bioimage Analysis Unit, Institut Pasteur, Paris, France [2] LLTech SAS, Paris, France [3] Department of Dermatology, Drexel University College of Medicine, Philadephia, PA

Motivation

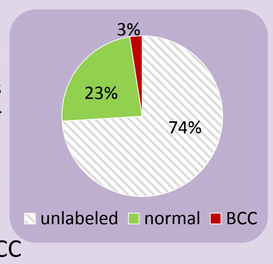
Skin cancer is the most common human malignancy, predominantly represented by **Basal Cell Carcinomas** (BCC). The gold standard treatment is **Mohs** surgery. It consists of successive removals and histological examinations of skin layers which guide further tissue extraction.

Our work aims to speed up the procedure by using a **non-invasive optical slicing** modality – **Full Field Optical Coherence Tomography** (FFOCT), together with an **automated diagnosis** of the cancerous areas – **Convolutional Neural Networks** (CNN).

This would lead to improved patient comfort and physician throughput.

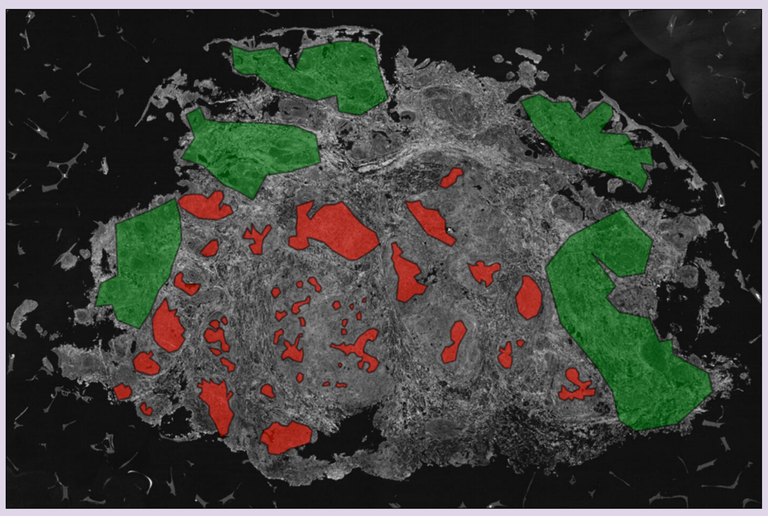
Data set

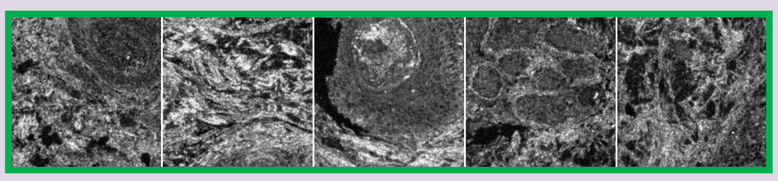
- 40 FFOCT images of tissue excisions
- 10 images contain cancerous areas
- 2 classes: **normal** and **BCC**
- images are manually labeled and diagnosed by a dermatopathologist, validated with the H&E frozen sections of samples
- 74% of total imaged area is unlabeled: background and abnormal tissue (sometimes surrounds BCC, morphologically irrelevant to either class, should consider separately)
- 23% is normal skin; only 3% is tumoral → class imbalance problem
- 108.082 patches = 59.112 normal + 48.970 BCC
- 80% train set, 20% test set

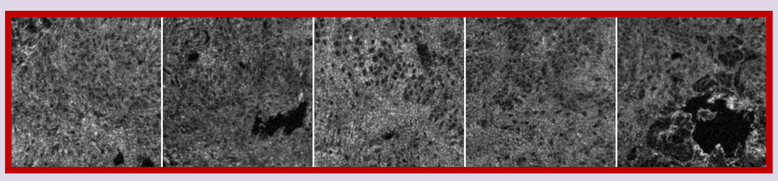


Data sampling & processing

- 16-bit DICOM (10 to 12 bits used) → 8-bit JPEG
- strong **speckle** noise → 3x3 Gaussian filter for fast smoothing while preserving the structures
- images split into **256x256px patches** – tradeoff between context capture and computational resources needed for CNN training
- class imbalance correction through patch **oversampling** with different step values : **170px** for normal class / **40px** for BCC
- data standardization (zero centering + normalization): robustness to variation in acquisition conditions
- data augmentation: flips, rotations, shifts





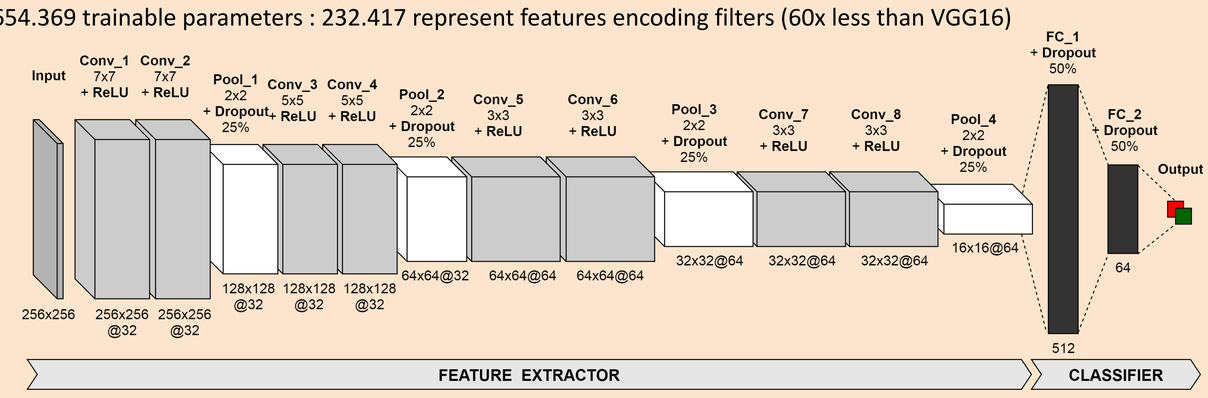


An annotated sample (11.808 x 8.352) and some patches extracted from it:
- **normal** patches with skin structures like collagen, hair follicles, sebaceous glands, sweat glands
- **BCC** patches with dense cancerous cell nuclei distributions and retraction artifacts

CNN

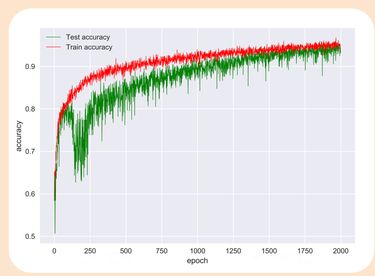
Architecture:

- 10 layers : 4 convolutional blocks (x2 layers each) + 2 fully connected layers
- 8.654.369 trainable parameters : 232.417 represent features encoding filters (60x less than VGG16)



Training:

- weights initialized with *Glorot*[4] method : gradients of each layer follow the same initial distribution; better and faster convergence
- minimize binary cross entropy loss with *Adam*[5] gradient descent on mini-batches of 40 samples over 2.000 epochs
- *Adam* (Adaptive Moment Estimation) : adapts the learning rate (step of descent) for each parameter according to its update frequency (e.g. larger updates for parameters which rarely change and vice-versa) while adding momentum for an accelerated optimization
- class weighting with respect to class ratio - 1÷1,2 - higher penalization for misclassifying BCC
- trained in one day (25hrs17min, 45sec per epoch) on 4 *Nvidia Tesla P100* GPUs
- coded using *Keras* *[github.com/fchollet/keras]* with *Tensorflow* *[tensorflow.org]* back-end



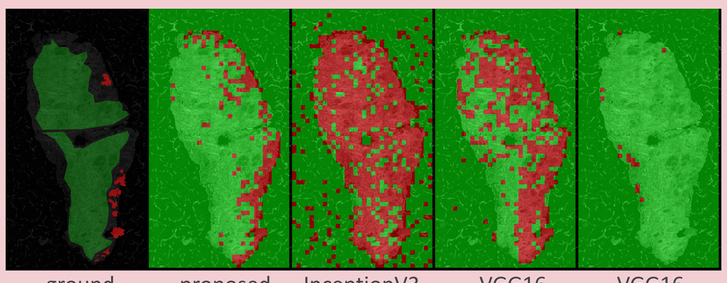
accuracy evolution during training

Results

Proposed CNN: prediction accuracy = **95,93%**

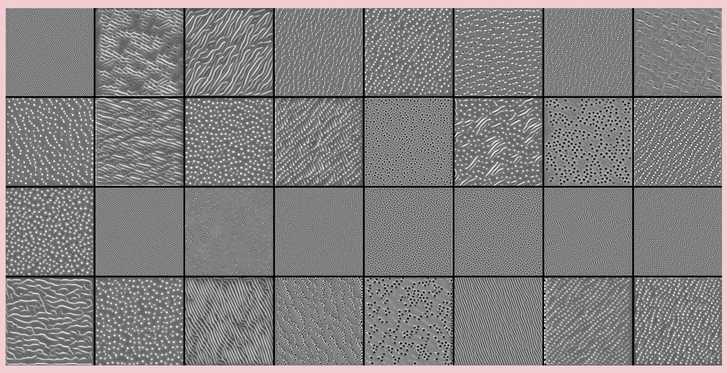
Out of the box models:

- *VGG16*[6] pre-trained on *ImageNet* : *acc* = 89,30%
- *InceptionV3*[7] pre-trained on *ImageNet* : acc = 90,79%
- early overfitting → architectures too complex cause data over specification → simpler architecture, trained from scratch


ground truth | proposed | InceptionV3 pre-trained | VGG16 pre-trained | VGG16 fine-tuned

Discussion:

- **visualize textures** which are learned by the filters (gradient ascent in the input space with respect to the filter activation loss → simulated input maximizing the activation of neurons forming that filter)
- filters could encode **distributions of cell nuclei** and **orientations of collagen fibers**
- unlabeled abnormal tissue predominantly classified as BCC


simulated inputs that would maximize the activation of **Conv_3** layer

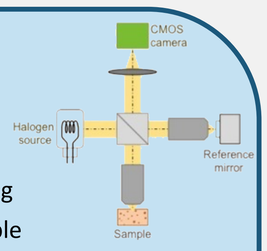
Conclusions

- ✓ **open promising research direction**: analyzing Full Field OCT images through deep learning
- ✓ **ease the integration** of a novel optical biopsy technology in the clinical environment by assisting pathologists with computer aided diagnosis (CAD) tools
- ✓ **reduce the costs and duration** of certain medical procedures, like Mohs surgery

Future work:

- ❑ collect more data
- ❑ consider new class : abnormal tissue
- ❑ understand the pathologist's diagnosis strategies and decision tree → adapt algorithms accordingly
- ❑ adopt a multi-scale approach: capture larger context, extract different information at different levels of zooming
- ❑ "unbox" the black box - understand the reasoning of the CNN → gain knowledge about the data
- ❑ Dynamic Cell Imaging (DCI) FFOCT[8] : morphologic + metabolic information on cells

LLTech Light-CT™ Scanner



"en face" optical slicing interferometry principle
resolution: **1 µm³** (intracellular)
penetration depth: **200 µm** to **1 mm**
speed of acquisition: **1 cm² / min**
size: 310 x 360 x 700 mm (portable)

References

1. E Dalimier and D Salomon, *Full-field optical coherence tomography: a new technology for 3d high-resolution skin imaging*, Dermatology, 2012.
2. G Litjens et al., *A survey on deep learning in medical image analysis*, Medical Image Analysis, 2017.
3. A Esteva et al., *Dermatologist-level classification of skin cancer with deep neural networks*, Nature Letter, 2017.
4. X Glorot and Y Bengio, *Understanding the difficulty of training deep feedforward neural networks*, PMLR, 2010.
5. DP Kingma and J Ba, *Adam: A method for stochastic optimization*, ICLR, 2015.
6. K Simonyan and A Zisserman, *Very deep convolutional networks for large-scale image recognition*, arXiv, 2014.
7. C Szegedy et al., *Rethinking the inception architecture for computer vision*, CVPR, 2016.
8. C Apelian et al., *Dynamic full field optical coherence tomography: subcellular metabolic contrast revealed in tissues by interferometric signals temporal analysis*, Biomedical Optics Express, 2016.

www.lltechimaging.com