

Medical Imaging with Deep Learning (MIDL 2018) Conference: Exploring Rejected Extended Abstracts

Nowadays, everybody and their grandmothers (and grandfathers) are utilizing deep learning for medical imaging problems. This has now become a [hot industry](#). Even though major computer vision (CVPR etc.) and medical imaging conferences (SPIE, ISBI, MICCAI) are currently dominated by deep learning methods, a conference specific to [deep learning applied to medical imaging](#) has just emerged in 2018.

We are lucky to have free access to videos and slides from latest research [here](#) :-)

This post covers the extended abstracts track. There is also a full conference track with lots of [rejected papers](#) to look into..but that is for another rainy day...

Judging by the reviews (that were apparently based on 2-3 page abstracts) shared openly, the selection process was competitive. According to openreview.net site(*), 35 papers were accepted and 44 papers were rejected (around 44% acceptance rate):

<https://openreview.net/group?id=MIDL.amsterdam/2018/Abstract>

I quickly looked into what was **rejected**. There are certainly interesting things to learn from research that did not quite make it in this round. Couple things stood out for me:

- **Out of scope: Annotation Tools:** You need to actually train a network. Although it is a necessity to have nice annotation tools in most cases, ["Building a mass online annotation tool for dental radiographic imagery"](#) was out of scope because it did not talk about any actual deep learning on the problem. Nevertheless, it does seem to allow for automated validation of model predictions (CNNs trained elsewhere) against expert annotations. I liked the idea from an end-to-end pipeline standpoint.
- **Out of scope: Computational improvements:** [PZnet](#) is a CPU-only platform for CNNs where you might have a huge CPU cluster and want to put it to good use. Alas, it was not too exciting for this conference.
- **Transfer learning is not hot anymore:** None of the accepted papers have transfer learning in their titles. All four papers ([1: Transfer Learning with Human Corneal Tissues: An Analysis of Optimal Cut-Off Layer](#), [2: Exploring Transfer Learning. Fine-tuning of Thyroid Ultrasound Images](#), [3: Characterizing Patterns in Phase Contrast X-Ray Computed Tomography Images of the Patellar Cartilage with Deep Transfer Learning](#), [4: How transferable are the deep features from false positive reduction network for lung nodule detection in CT to malignancy prediction](#)) on this topic were rejected. Mostly due to small sample size or not using a novel architecture.
- **Not novel:** Basically all papers tackled important medical image analysis problems. Novelty seems to be assessed mostly in terms of the proposed architecture. For instance, it appears that end-to-end solutions for clinical environments were not favored here unless there is a novel architecture proposed in the paper. [A Lightweight ConvNet for 4D Multi-Structure Segmentation of Cardiac Cine-MRI](#) was rejected due to being an U-NET like architecture and that the problem domain was already covered extensively in

Kaggle and MICCAI competitions. The dashboard by itself apparently was not too exciting for this conference. Some other papers that were rejected based on lack of architecture novelty or lack of comparisons with other architectures: [laser ophthalmoscopy](#), [segmentation of cervical vertebrae on CT images](#), [tooth detection on radiographs](#), [detection of midline brain abnormalities](#), [U-net for artery-vein segmentation in retinal images](#), [peri-tubular capillary counting](#), [edema segmentation](#), [Detecting landmarks on MR and Camera Images](#), [It doesn't take a whole U-Net to find brain tumors](#), [Multi-label 4-chamber segmentation of echocardiograms](#), [diffuse interstitial lung disease using U-Net](#), [Projection Data Upsampling for Sparse and Low-Dose CT Scout Scans](#), [Detecting Local Infiltration of Lung Cancer Primary Lesions on Baseline FDG-PET/CT](#), [Mammogram classification in BIRADS standard](#), [Optical Coherence Tomography Images in Neurodegenerative Disease](#)

- **Maybe too novel:** ["Spectral Analysis Towards Geometric Auto-Encoding of Subcortical Structures"](#) is an example from the emerging field of geometric deep learning. Alas it was found to be too difficult to follow by one of the [reviewers](#).
- **Too preliminary:** There are some interesting ideas here but mostly half cooked. ["Visual interpretability for patch-based classification of breast cancer histopathology images"](#) tackles the interpretability of network decisions, which is an important topic. Interesting work but not enough experimental evidence was given to convince the reviewers. Another example is on [Meta-Learning for Medical Image Classification](#). This one is proposing a work-around for small sample issue with transfer learning from a large dataset such as ImageNet. The authors claim that "meta-learning to obtain the initial weights for further fine-tuning can provide better results than the current state-of-the-art methodology". It seems to beat the baseline but not more sophisticated deeper networks applied to the same Kaggle dataset. [Personalised Patient Embeddings - Towards a Whole Data Health Profile](#) seems to be about multi-modal deep learning including images, text, genetic information and blood results per patient. It proposes Sliced-wasserstein autoencoder(SWAE) or the Deep Copula Information Bottleneck models. Both the problem and proposed solution were very vaguely described. It is not clear if any actual implementation has been developed by the authors.
- **Double dipping:** ["Using deep learning to develop a fully automated, real-time 3D-ultrasound segmentation tool to estimate placental volume in the first trimester"](#). Very cool clinical application using 3D U-NET; however it seems that since the work already has been accepted elsewhere, they could not reveal any useful information in their abstract (not even a figure) for review here.
- **Lack of Details vs. Limited space:** I also noticed that in some cases, the reviewers were complaining ([1: A multi-level deep learning algorithm to estimate tumor content and cellularity of prostate cancer](#), [2: Tissue segmentation in volumetric laser endomicroscopy data using U-net and a domain-specific loss function](#), [3: Accuracy enhancement of CT kernel conversions using convolutional neural net for super-resolution with Squeeze-and-Excitation blocks and progressive learning among smooth and sharp kernels](#), [4: Residual CNN-based Image Super-Resolution for CT Slice Thickness Reduction using Paired CT Scans](#)) about lack of details as a basis for rejection, while the authors were [complaining](#) about lack of enough space to provide them. I also think packaging everything in a 2-3 page abstract along with references is a little bit difficult.

- **Super-resolution.** This must be one of the popular topics here. Interestingly all three abstracts mentioning super-resolution on their titles were rejected but there are a couple presented in the [conference](#) track. The goal here is to use deep learning to come up with a way to create high-resolution images from low-resolution images. Here two papers ([Squeeze-and-Excitation blocks and progressive learning among smooth and sharp kernels](#), [Residual CNN-based Image Super-Resolution for CT Slice Thickness Reduction](#).) aim to do this on CT images, while the other works on MRI images ([MRI Single Image In-Plane Super Resolution Using Mixed-Scale Sense CNN](#)) .

There were some controversial papers that received mixed reviews (some accept, some reject) but eventually rejected:

- **Deep learning & Atlas based segmentation hybrid:** This is one area of interest for me. As we all know, deep learning models are dumb and have no clue about what they are looking at. You can train the same model on cat images or cardiac images. It is a cool idea to try and [introduce anatomical knowledge](#) into these models. One reviewer clearly accepts the paper even though the validation was not done quantitatively, while the second reviewer is not impressed because "UNet and multi-atlas registration based segmentation is widely used".
- **Application of geometric deep learning to brain imaging.** [Spectral Analysis Towards Geometric Auto-Encoding of Subcortical Structures](#) is about statistical shape analysis. As one [reviewer puts it](#), this is a novel technique to apply deep learning to meshes instead of images directly, however, the details were not found to be clear enough to judge. It is based on multi-scale mesh based shape representation. The algorithm seems to be learning a layered representation of these shapes in terms of a hierarchy of shape signatures (based on spectral wavelet transforms). Learning is done layer-wise using unsupervised learning such as k-means or variational Bayes EM. In the past, I was obsessed with content based 3D shape retrieval. The notion of shape signatures was also used there in conjunction with similarity measures. This takes me back to those times.
- **Predicting follow up images to track disease progression .** The goal is to develop an unsupervised model to learn static anatomical structures and the dynamic changes of the morphology due to aging or disease progression. The paper [Unsupervised Representation Learning of Dynamic Retinal Image Changes by Predicting the Follow-up Image](#) uses almost "4000 OCT images of about 200 patients with macular degeneration who were scanned over 24 months". One reviewer found the experiments confusing.
- **Cycle-Consistent Generative Adversarial Networks for Image Segmentation.** This seems to be an interesting idea to perform [Epithelial tissue image segmentation using GANs](#). According to the authors: "The model consists of two generators, one that maps from the image to the segmentation domain and a second that maps from the segmentation to the image domain, and two associated adversarial discriminators. A so-called cycle-consistency loss regularizes the mapping and enforces a relationship between an image in the segmentation and the image domain". They find the performance of Cycle-GAN comparable to U-NET that had to be trained on a paired input image-segmentation dataset. As the reviewers point out they only run experiments

with the fully annotated training set, therefore did not sufficiently show that it works when we lack annotations. The second paper using the same general idea is on [liver lesion segmentation](#). It proposes an improved U-Net architecture called the polyphase U-Net as a generator in the Cycle-GAN. However, the reviewers pointed out that the performance did not match up to the state-of-the-art in the competition where the dataset came from.

- **Deep Learning-Based 3D Freehand Ultrasound Reconstruction.** Even though ultrasound imaging is cheap and safe, it is challenging to construct 3D volumes without external probes or tracking hardware. According to one reviewer, the authors have previously presented a deep learning based method to build 3D volumes from a series of 2D images at MICCAI. [In their MIDL submission](#) they have added an inertial measurement unit (IMU) hardware to model more realistic operator hand motions via a gyroscope. It must indeed be challenging to do this solely based on images after all. However, it seems this was no longer novel enough for the reviewers.
- **Reconstruction of sparsely sampled Magnetic Resonance Imaging measurements with a convolutional neural network.** [This one](#) talks about Compressed Sensing accelerated Magnetic Resonance Imaging (MRI) and how a neural network can be used to decode accelerated, undersampled MR acquisitions, eliminating the need for reconstruction algorithms. One reviewer likes it while the other bashes it questioning methodological novelty and asking for comparisons with other architectures.

There were some interesting ideas (or innovations one might say) borrowed from previous literature but did not quite make it in the end:

- **Improved Deep Learning Model for Bone Age Assessment using Triplet Ranking Loss** Strangely [this paper](#) received marginally above threshold ratings from both reviewers but ended up being rejected in the end. It uses the idea of regularizing feature embeddings with respect to clustering of similar cases together in the feature space by introducing ranking in the loss function. I guess this triplet ranking loss concept was proposed by [Google in 2015](#) for face recognition problems. To my understanding, the reviewers were saying that the experiments on the 3-stage architecture were not clearly explained or had some issues, so they were not convinced.
- **Curriculum learning from patch to entire image for screening pulmonary abnormal patterns in chest-PA X-ray** This paper borrows the notion of [curriculum learning from Bengio et al.'s 2009 ICML paper](#) that proposes training gradually from simple to more complex concepts. In this paper the authors first train a RESNET 50 using Imagenet dataset, then train it using lesion extracted patches. Then they fine-tune it using the entire images. Their idea is that it would be difficult to train the network on the more complex whole images where there are other organs etc. exist. Interesting idea. However, the reviewers had issues with clarity of the writing and lack of some details.
- **Detection of Gastric Cancer from Histopathological Image using Deep Learning with Weak Label** [Weak supervision](#) is a work-around to handle lack of high quality annotated datasets from domain experts where large scale albeit noisy and lower quality annotations are gathered using cheap annotators or programmatically. Apparently the idea is not new in machine learning but has been gaining popularity in deep learning recently. The goal is to predict slide level labels, however since the images are large, it

uses a patch based approach where the algorithm ultimately generates a probability map by stitching patch level predictions. Then a random forest algorithm predicts the label for the whole slide using the probability maps.

This paper uses slide-level weak labels if there are no region-level labels available for the patches. It was not really found novel by the reviewers I am guessing.

This conclusion is not surprising at all, but here it goes: based on common reviewer comments, it seems that having a large dataset is preferable and that is becoming less of an issue with lots of free datasets being available via competitions etc. One rule of thumb however, is to make sure that the method is at least as good as the state-of-the-art if one is using a competition/challenge dataset. Even if the dataset/application domain is novel, some innovation on the architecture along with comparison to established architectures is expected. Of course, in all cases, it is essential to be able to clearly convey all relevant information within the limited number of pages (2 or 3).

*Note: authors are allowed to remove their papers from the site, therefore these numbers might not be accurate.

Disclaimer: the opinions stated in this blog post are my own. They do not represent my employer's views or opinions on related topics.