# Image analysis in medical applications

Anna Breger, MSc BSc

Department of Mathematics, University of Vienna
Supervisor: Assoc. Prof. Martin Ehler

Laboratory of Mathematics in Imaging (LMI),
Brigham and Women's Hospital, Harvard Medical School
Supervisor: Dr. Carl-Fredrik Westin

Scientific report Marshall Plan Scholarship
February 25 - May 26, 2019

# Overview

At the beginning of my stay at the Laboratory of Mathematics in Imaging (LMI) in Boston I was invited to give a talk at the LMI and the Applied Chest Imaging Laboratory to present my research. Moreover, I could attend several seminars in the LMI itself and related research groups, including the Applied Chest Imaging Laboratory, the Golby Lab at Brigham and Women's Hospital and the Computer Science and Artificial Intelligence Laboratory (Csail) at MIT. It was a perfect possibility to connect with a wide range of researchers and find suitable people for collaborations. That yielded two main projects during my stay:

1. Coil combination with orthogonal projections

2. Brain tissue segmentation with an augmented target loss function

In both projects I could merge different parts of my PhD project with applications in Magnetic Resonance Imaging (MRI). At the end of my stay I presented a poster about my research on MRI reconstruction with orthogonal projections at the 7th Annual Radiology Research Symposium, Brigham and Women's Hospital.

The outline of the report is as follows: in the first chapter I will present my work on orthogonal projections for coil combination in the form of a preprint paper with the status of September 2019. The second chapter consists of three parts: first, a section from my paper [1] introduces the method of augmented target loss functions, then I present results of the application to brain tissue segmentation with MRI data, and at the end the application to photoreceptor segmentation in optical coherence tomography (OCT) data. All projects are ongoing and shall yield more research results in the future.

# Contents

# Chapter 1

# Orthogonal projections for coil combination of MRI data

# On the reconstruction accuracy of multi-coil MRI with orthogonal projections for coil combination

Anna Breger[1], Gabriel Ramos Llorden[2], Gonzalo Vegas Sanchez – Ferrero[3], W. Scott Hoge[3], Martin Ehler[1], and Carl-Fredrik Westin[4]

[1]Department of Mathematics, University of Vienna, Austria
[2]Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[3]Department of Radiology, Brigham and Women's Hospital, Boston, MA
[4]Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

September 2019

**Abstract**

MRI signal acquisition with multiple coils in a phased array is nowadays a standard setting. It increases the signal-to-noise ratio (SNR) and can be accelerated by using parallel imaging methods. Some of these methods like GRAPPA or SPIRiT yield fully sampled data in the k-space, which need to be combined to get the final images. For this last step of the image reconstruction, often the root sum of squares (rSOS) is used. This straightforward method works well for coil images with high SNR, but can yield problems in images with artifacts or low SNR in all individual coils. We aim to analyze that final coil combination step in the framework of linear compression. With two data sets, a simulated and an in-vivo, we use random projections as a representation of the whole space of orthogonal projections. This allows us to study the impact of incorporation of linear compression on image space data, including principal component analysis (PCA) that provides an orthogonal projection. The $L_2$ error, variance, SNR and visual results serve as performance measures to describe the final accuracy of reconstructions. We study their relationships and observe that the minimal $L_2$ error does not necessarily correspond to the best visual results. In terms of visual evaluation and SNR, the incorporation of compression with PCA outperforms all other methods, including rSOS on the uncompressed image space data.

## 1 Introduction

Magnetic Resonance Imaging (MRI) is a unique medical imaging modality that provides excellent image quality without ionizing radiation, but on the other

1

hand is relatively slow. The acquired samples, also known as the k-space data, are samples of the Fourier transform of the MR image. To reconstruct an accurate MR image, sampling theory indicates the total number of k-space data that must be acquired to avoid artifact-like reconstruction. As this number is relatively large and cannot be arbitrarily reduced, the total scan time cannot be shortened without compromising the image quality.

A turning point in MRI reconstruction was the implementation of phased-array coils in the acquisition pipeline. With phased-array coils, the k-space data is acquired in multiple receivers/coils simultaneously. On top of increasing the signal-to-noise (SNR) ratio, a phased-array coil acquisition allows accelerating the total acquisition time [1]. These methods, known as Parallel Imaging methods, exploit the fact that there exist complementary k-space data information in each coil/receiver. The k-space data set is undersampled, but the multiple measurements allow the reconstruction of missing k-space samples in a kind of inverse problem reconstruction framework. See [2] for an overview of basic reconstruction algorithms for parallel imaging and their history.

The computational costs and required memory of the reconstruction algorithms highly depend on the dimension of the phased array, i.e. the number of receiver coils. Several coil compression methods have been developed that reduce to a smaller set of virtual channels without a significant loss of SNR, see e.g. [3],[4],[5]. Moreover, PCA-based methods have shown to even have a beneficial denoising effect, e.g. [6]. In [7], the optimal linear projection for coil compression based on the resulting SNR is derived. Note that this is related to our analysis approach, but we work in the image space instead of the k-space.

After the obligatory compression step, PI reconstruction methods are applied and can result in different output spaces. Some of the methods reconstruct the image directly (image-based methods), e.g. SENSE [8]. K-space based methods retrieve the missing k-space data, thereby providing a fully-sampled k-space array per each receiver/coil, e.g. GRAPPA or SPIRiT [9, 10]. Next, all k-space data is inversely Fourier transformed and the resulting image array has to be combined into a single final image. This image represents the conventional MR image and is used for analysis, quantification, or visualization purposes.

In [1], optimal methods to combine the arrays from phased array elements have been developed. These methods rely on detailed knowledge of each coil's magnetic fields. In practice, the exact position of the coils is not always known or not possible to compute, because of computational limitations. To overcome this issue, a method that combines the data without detailed knowledge of the coils, while preserving a high SNR, is needed. For this purpose, the root sum of squares (rSOS) has become the standard method of combining multi-coil images in MRI [1], [11]. For arrays with high SNR the rSOS yields nearly optimal reconstruction, whereas problems arise if all coils yield low SNR. Especially data with artifacts are problematic since rSOS weights them equally to the non-defective parts.

Noise has a huge impact in MRI imaging and is often caused by thermal noise in the reception and the patient's body itself due to radiofrequency emissions. Moreover, the SNR in the coil images depends highly on the scanner, imaging

2

modality (e.g. T1/T2/diffusion weighted) and the number of coils/receivers. Extensive statistical noise analysis can be found in [12].

In this work, we aim to analyze the coil combination in a comprehensive way by comparing different performance measures. To address coil combination independent of prior knowledge, we will study the rSOS in the framework of linear image space compression via orthogonal projections. Random projections will serve us as a tool to study the correlation between reconstruction error and voxel variance of the varying reconstructions. Correlation analysis with random orthogonal projections has been studied in a related context in [13], yielding an underlying understanding of the relation between important information preservation features in data combination and compression. Moreover, we will describe the performances regarding the SNR and visualization, indicating a clear improvement by compressing the image space data with PCA.

The outline is as follows: first we will explain how rSOS can be interpreted in the context of orthogonal projections. Moreover, we will describe samples of random orthogonal projections, that shall serve us as coverings enabling numerical experiments. Then we describe the two data sets, a simulated and an in-vivo, that we use for experimental investigations. The simulated data set allows us to compare the behavior with different noise levels in the coils. In the results, we show scatter plots describing the relation between the $L_2$ reconstruction error and the voxel variance in reconstructed magnitude images. Finally, we will interpret and discuss the results in Section 5.

## 2    Methods

We aim to study the impact on reconstruction accuracy by incorporating linear compression in the last coil combination step with rSOS. For example, phased-array data that has been processed with GRAPPA or SPIRiT in k-space, needs to be combined as the final step of the reconstruction pipeline. Before combining this fully sampled, multidimensional image space data, we include linear compression with orthogonal projections.

The common reconstruction pipeline including our linear compression can be summarized as follows:

$$\hat{y}_i \in \mathbb{C}^d \xrightarrow{\text{GRAPPA}} \hat{x}_i \in \mathbb{C}^d \xrightarrow{\text{IFFT + abs}} x_i \in \mathbb{R}^d \xrightarrow{\text{projection}} px_i \in \mathbb{R}^k \xrightarrow{\text{rSOS}} \|px_i\|_2 \in \mathbb{R}$$

where $\hat{y}_i$ corresponds to an undersampled k-space voxel, $\hat{x}_i$ is fully sampled in k-space after some PI reconstruction (e.g. GRAPPA) and $x_i$ is the image space data. Our analysis takes place on the image space data in $\mathbb{R}^d$ that is subsequently projected to $\mathbb{R}^k$ with $k < d$.

### 2.1    Framework of orthogonal projections

As the basis for our analysis we will use random orthogonal projections yielding different linear coil compressions. We will work with image data consisting of $d$ channels that are combined into one final magnitude image. To do so, we study

3

the space of $k$-dim linear subspaces of $\mathbb{R}^d$, which can be identified by orthogonal projections

$$\mathcal{G}_{k,d} = \{p \in \mathbb{R}^{d \times d} : p^2 = p, p^T = p, \mathrm{rank}(p) = k\}, \tag{1}$$

called the Grassmannian manifold.

Let $x = \{x_i\}_{i=1}^m \in \mathbb{R}^d$ be an image space data set with $m$ voxels measured by $d$ coils. Then, the final image volume is given by $\|px\|_2$ with $p$ in $\mathcal{G}_{k,d}$, where $d$ is the fixed number of channels and $k < d$ varies. This corresponds to projecting the $d$ coil channels to different dimensions $k$ and computing root sum of squares (rSOS) afterwards. Note that we can study the rSOS itself in this context, since $\mathrm{rSOS}(x) = \|x\|_2$ and it holds for all $p \in \mathcal{G}_{k,d}$ that the expectation value

$$\mathbb{E}[\|px_i\|_2] = c \cdot \|x_i\|_2 \quad \forall x_i \in \mathbb{R}^d, \tag{2}$$

with $c = \left(\Gamma(\frac{k+1}{2})\Gamma(\frac{d}{2})\right)/\left(\Gamma(\frac{k}{2})\Gamma(\frac{d+1}{2})\right)$, where $\Gamma$ denotes the Gamma function.

This is based on the Chi distribution and the fact that the length of a random unit vector projected onto a fixed $k$-dimensional subspace has the same distribution as the length of a unit vector in $\mathbb{R}^d$ being projected onto a random $k$-dimensional subspace (see e.g. [14]).

**Remark 2.1** *The equality (2) allows us to analyze the rSOS itself in the framework of orthogonal projections, i.e. no added coil compression in the image space. Summing up the projected voxels obtained by a reasonably big sample set of random orthogonal projections $p = \{p_l\}_{l=1}^n \in \mathcal{G}_{k,d}$, yields approximately the rSOS combined voxel up to the constant c, i.e.*

$$\frac{1}{n} \sum_{l=1}^n \|p_l x_i\|_2 \approx c \cdot \|x_i\|_2. \tag{3}$$

*Since we linearly rescale the final image volumes between [0,1] to enable error estimation with the ground truth, the constant is negligible. Note that also principal component analysis (PCA) yields an orthogonal projection that lays in the Grassmannian manifold and therefore can be studied in that context.*

We will use random orthogonal projections, i.e. projections $p \in \mathcal{G}_{k,d}$ distributed according to the orthogonally invariant probability measure $\mu_{k,d}$, as samples of orthogonal projections, see e.g. [15], [16]. The $L_2$ error will serve us as a measure of reconstruction accuracy for combining a image space data set $x = \{x_i\}_{i=1}^m \in \mathbb{R}^d$. Note that w.l.o.g. $x$ contains here just the voxels in the region of interest and not the full image volume with background. For a fixed projection $p \in \mathcal{G}_{k,d}$, the error between some provided ground truth $y = \{y_i\}_{i=1}^m$ and the combined magnitude image voxels $\|px\|_2 := \{\|px_i\|_2\}_{i=1}^m$, is then given by

$$\mathrm{Err}(y, \|px\|_2) := \frac{1}{m} \sum_{i=1}^m \left(y_i - \|px_i\|_2\right)^2. \tag{4}$$

4

We aim to study the relation between the reconstruction error and the variance in the reconstructed magnitude images, which can be interpreted as contrast in noise-free images. The variance of the final magnitude images depending on the projection method can be computed as

$$\text{Var}\left(\|px\|_2\right) = \frac{1}{m(m-1)} \sum_{i<j} \left(\|px_i\|_2 - \|px_j\|_2\right)^2. \tag{5}$$

Moreover, we will use the mean signal-to-noise ratio to interpret the performance. Note that in MRI it is common to use the non-squared version, i.e.

$$\text{SNR}\left(\|px\|_2\right) = \frac{1}{m} \sum_{i=1}^{m} \frac{\|px_i\|_2}{\sigma}, \tag{6}$$

where $\sigma$ corresponds to estimated standard deviation of the noise, see e.g. [12].

## 2.2 Random projections as coverings

To enable a numerical analysis, we need a finite set of orthogonal projections that represents the overall space well, i.e. covers the Grassmannian $G_{k,d}$ properly. To measure how well a set covers the underlying space, we use the definition of the covering radius.

**Definition 2.2** *Let the* covering radius *of a finite set* $\{p_1, \ldots, p_n\} \subset \mathcal{G}_{k,d}$ *be denoted by*

$$\rho(\{p_l\}_{l=1}^n) := \sup_{p \in \mathcal{G}_{k,d}} \min_{1 \leq l \leq n} \|p - p_l\|_{\mathrm{F}}, \tag{7}$$

*where* $\|\cdot\|_{\mathrm{F}}$ *is the Frobenius norm.*

Note that the smaller the covering radius, the better the finite set of projections $\{p_l\}_{l=1}^n$ represents the entire space $\mathcal{G}_{k,d}$: it yields smaller holes and the points are better spread.

Let $\mu_{k,d}$ denote the normalized Riemannian measure on $\mathcal{G}_{k,d}$ as before. According to [17], the expectation of the covering radius $\rho$ of $n$ random points $\{p_j\}_{j=1}^n$, independent identically distributed according to $\mu_{k,d}$, satisfies [1]

$$\mathbb{E}\rho \sim n^{-\frac{1}{k(d-k)}} \log(n)^{\frac{1}{k(d-k)}}. \tag{8}$$

Following the definition of asymptotically optimal covering in [16], the expectation yields an optimal covering radius up to a logarithmic factor $\log(n)^{\frac{1}{k(d-k)}}$. To remain flexible in the dimension of $\mathcal{G}_{k,d}$, i.e. the choice of $k$ and $d$, and the number of projections $n$, we will work here with random projections distributed according to $\mu_{k,d}$ rather than constructing optimal covering sequences as in [16]. These random orthogonal projections can be efficiently computed by the $QR$

---

[1]We use the symbol $\sim$ to indicate that the corresponding equalities hold up to a positive constant factor on the respective right-hand side.

decomposition of a matrix $M = QR$ with independent identically distributed entries [18].

In the following we will use finitely many samples of random projections in the experimental analysis.

# 3   Data

We will run our numerical analysis on two different MR data sets: a simulated T1-weighted data set from brainweb ([19], [20], [21]) and an in-vivo data set from a head coil receiver.

## 3.1   Simulated data set

Simulation experiments were conducted to assess the quality in image reconstruction for different type of projections in a controlled, rigorous manner. To do so, first, a ground-truth volume was created with the popular numerical simulator BrainWeb [22]. A (magnitude) multi-slice T1-weighted volume was simulated with a Spoiled Fast Low Angle Shot (SFLASH) sequence with the following parameters: TR/TE = 20 /10 ms, flip angle of 90 degrees and ETL = 1. Matrix size: $181 \times 21 \times 76$ with isotropic voxel size of 1 mm.

Next, simulated images with 32 channels, mimicking a 32-channel coil acquisition, were created. First, synthetic coil sensitivity profiles were simulated assuming a smooth Gaussian profile [12]. Voxel-wise multiplication of those coil sensitivities by the simulated ground-truth image creates the 32 coil-based images. Uncorrelated complex Gaussian noise with zero mean and standard deviation $\sigma$ was added. Finally, to simulate a magnitude-based acquisition, the absolute value of 32 noisy images were taken. The value of $\sigma$ was chosen differently to recreate experiments with different noise levels, resulting in the SNR values $2, 3, 5.5$ and $11$ for the mean rSOS reconstruction.

## 3.2   In-vivo data

MR data was acquired in-vivo from a healthy volunteer using a Siemens (Erlangen, Germany) 3T Prisma equipped with a 32-channel head coil receiver. An Echo-Planar diffusion sequence was employed to acquire 24 slices in a 2mm isotropic volume, with slice-thickness 2mm (TR=3.2 sec, TE=85ms, flip-angle=90, matrix size 128x128, FOV: 256mm by 256mm). A T2-weighted volume was acquired, with no diffusion weighting (a "b=0" image), followed by 62 volumes acquired using a single repeated diffusion vector and diffusion setting of b=1000. The measured EPI data was reconstructed using Dual-Polarity GRAPPA (DPG) to minimize Nyquist ghosts, see [23]. In-plane acceleration of the original data was $R = 2$. The ground-truth image was formed by motion-correcting the 62-volume diffusion-weighted series, using the Advanced Normalization Tools (ANTs) library ([24]), and averaging across time. The first diffusion direction was studied regarding linear coil combination and compared

to the computed the ground-truth. Few outliers have been removed by using the 99.99th percentile and setting the remaining 0.01% to the maximum of the used percentile.

# 4   Results

Both studied image volumes have a phased-array with 32 channels, therefore we work with projections $p$ in $G_{k,d2}$ with $d = 32$ and varying $k$. In the following scatter plots we see how the $L_2$ reconstruction error (4) relates to the voxel variance (5) and state corresponding mean SNR values in the visualization.

Each point in the plots corresponds to some linear compression in rSOS, yielding a final magnitude image volume as described in the previous sections. The symbols + correspond to a combined image by a projection $p \in G_{k,32}$, i.e. $\|px\|_2$, and the colors correspond to the different dimensions $k$:

$$+ \, p \in \mathcal{G}_{1,32}, + \, p \in \mathcal{G}_{4,32}, + \, p \in \mathcal{G}_{12,32}, + \, p \in \mathcal{G}_{20,32}, + \, p \in \mathcal{G}_{28,32}, + \, p \in \mathcal{G}_{31,32} \,.$$

As described in Section 2.2 the projections $p$ serve as a covering of the underlying space and are chosen randomly according to the orthogonally invariant probability measure $\mu_{k,d}$. For the simulated data set (see Section 3.1) $n = 500$ projections have been randomly chosen for every space $\mathcal{G}_{k,32}$ with $k \in \{1, 4, 12, 20, 28, 31\}$, for the in-vivo data set (see Section 3.2) $n = 1000$.

The symbol $\square$ corresponds to the rSOS method, i.e. $\|x\|_2$, and $\circ$ to the compression with the orthogonal projection provided by PCA. The colors correspond to the different spaces $\mathcal{G}_{k,32}$ as stated above.

Figure 1 contains the scatter plots for 4 different noise levels in the simulated data set and Figure 2 shows corresponding image cross-sections. Figure 3 shows the scatter plot and image cross-sections of the in-vivo data set.

# 5   Discussion

In Figure 1 and Figure 3 we illustrate the scatter plots of the described linear compression methods regarding the two data sets. To simplify the visualization we display the spaces $\mathcal{G}_{k,32}$ only for $k \in \{1, 4, 12, 20, 28, 31\}$. In all plots we can see that the smaller the dimension $k$, the more the projections are spread regarding $\mathrm{Err}(y, \|px\|_2)$ and $\mathrm{Var}\left(\|px\|_2\right)$; the reconstructions including random projections from $G_{1,32}$ are widely distributed, whereas using the projections in $G_{31,32}$ are clustered closely around the rSOS. The smaller the $k$, the more original information can be randomly dismissed: the preservation and loss of original information varies less for image space data compressed with a random projection $p \in G_{31,32}$ than for images space data compressed with $p \in G_{1,32}$.

For lower noise levels in the simulated data, Figure 1 (a)-(b), we can directly see that the correlation between the $L_2$ reconstruction error and the variance within the final combined images is very strong. Since variance can be interpreted as contrast in images with high SNR, it shows that high contrast directly

7

(a) SNR = 10.64
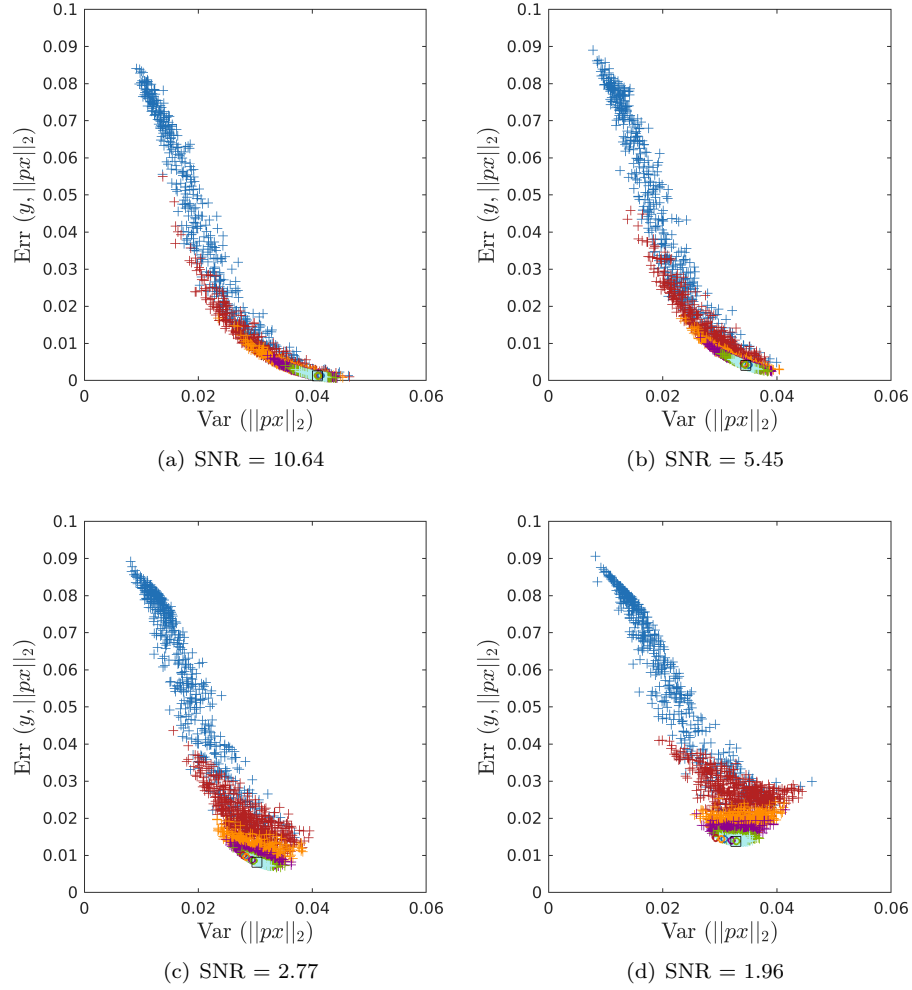
(b) SNR = 5.45

(c) SNR = 2.77

(d) SNR = 1.96

Figure 1: Scatter plots for the simulated data set with different noise levels, showing the reconstruction error (4) and variance (5) obtained by combining the coils with random projections in $\mathcal{G}_{k,32}$, PCA and rSOS. A varying amount of Gaussian noise was added in each coil assuming no correlation [12]. The SNR value corresponds to the mean over all voxels in the reconstructed rSOS image volume.
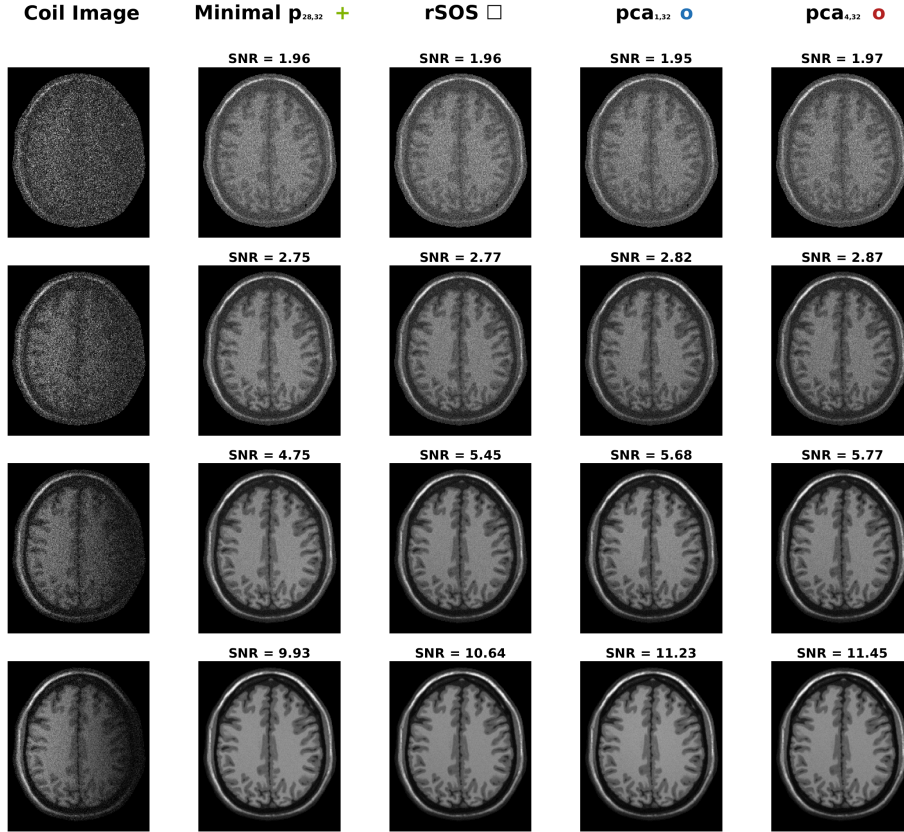
8

Figure 2: Cross-sectional image corresponding to the scatter plots in Figure 1 for the simulated data set with different noise levels. The left column shows the first channel of the simulated noisy 32-dim coil array before the coil combination step. The second column shows the final image obtained by a sample of random projections $p \in \mathcal{G}_{28,32}$ with the minimal reconstruction error (4). The third column shows the image provided by rSOS as coil combination method. The second last columns show the magnitude images when using PCA in $\mathcal{G}_{1,32}$ and $\mathcal{G}_{4,32}$ for compression, yielding the highest mean SNR.
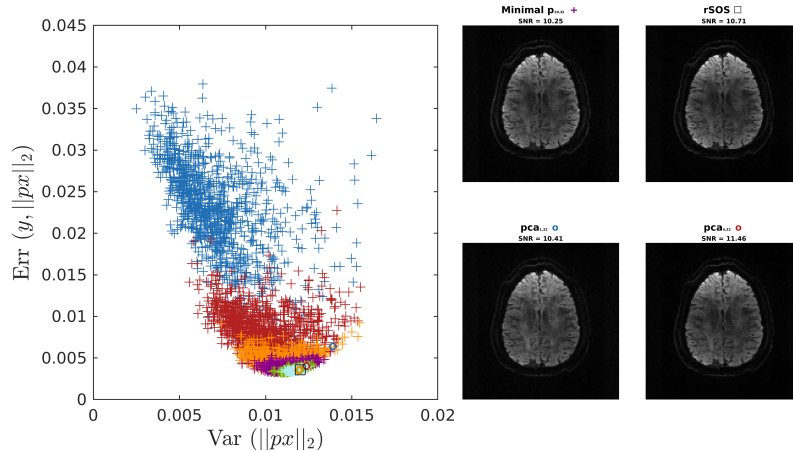
Figure 3: Left - Scatter plot for the in-vivo data set showing the reconstruction error (4) and variance (5) obtained by combining the coils with random projections in $\mathcal{G}_{k,32}$, PCA and rSOS. Right - Ground Truth and final images obtained by different coil combinations of a cross-sectional image example. The SNR value corresponds to the mean over all voxels in the reconstructed image volume, see (6).

relates here to a good reconstruction in the $L_2$ sense. However, the higher the noise level, the more noise is also described in the measured variance and therefore it cannot directly be interpreted as contrast any more. In the simulated data set, Figure 1 (c)-(d), this can be seen in the change of correlation behavior, where higher variance does not relate to a lower reconstruction error for all dimensions $k$. In Figure 1 (d) the projections from the spaces $G_{k,32}$ with $k = \{12, 20, 28, 31\}$ do not show any connection between the variance and $L_2$ error. This might happen because the high noise level in the original data influences the measure of variance strongly, overriding the underlying meaning of contrast. Interestingly, for $k = 1$ there is still some negative correlation, indicating that the randomization leads to several poor reconstructions where noise is secondary in comparison to issues with contrast. The scatter plot corresponding to the in-vivo data (Figure 3) shows similar behavior. Because of this varying behavior, the variance itself does not act as a useful measure of reconstruction performance in this experimental setup.

We can see that in all experiments, rSOS itself as well as including some PCA yield nearly optimal $L_2$ reconstruction error, but are always outperformed by some random projections. Nevertheless, regarding SNR and visualization, these reconstructed images with random compression cannot compete with PCA compression or rSOS. That indicates some contradicting behavior, when measuring the reconstruction performance with the $L_2$ error versus the SNR. Indeed, when using PCA for the simulated data set, the $L_2$ error is the lowest in $G_{31,32}$,

10

whereas highest SNR is always achieved by PCA in $G_{4,32}$, which contradictory yields the worst $L_2$ error. Also in the in-vivo data set the highest SNR was achieved by using PCA in $G_{4,32}$, which again does not correspond to the lowest $L_2$ error. Following the visual results it seems that the SNR describes here the visual performance better than the $L_2$ error. The compression with PCA in $G_{4,32}$ yields the highest SNR in all experiments and therefore outperforms rSOS consistently. Compression with the standard PCA in $G_{1,32}$ yields predominantly better results than rSOS, but yields an insufficient visual result on the in-vivo data.

# 6 Summary

Based on random orthogonal projections we have shown a numerical investigation on reconstruction accuracy regarding rSOS coil combination with linear compression. Two different MR data sets were used for our experiments; a simulated T1 weighted data set with varying amount of noise and an in-vivo diffusion weighted data set. We used diverse measures of performance to evaluated the accuracy of the reconstructions. For the lower noise levels in the simulated data, the $L_2$ reconstruction error yields strong correlation with the variance, but the behavior changes for higher noise levels and the in-vivo data. Moreover, measuring $L_2$ error and SNR acts contradictory in terms of optimality and in these cases we observe that the visual evaluation corresponds more to the SNR. The highest SNR values were achieved by incorporating PCA as compression before using rSOS, outperforming rSOS with no compression in all experiments. This clearly suggests to use PCA on image space data before computing the final coil combination with rSOS, yielding a beneficial denoising effect with higher SNR.

# 7 Acknowledgment

# References

[1] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller. The NMR phased array. *Magnetic Resonance in Medicine*, 16(2):192–225, 1990.

[2] Mark Griswold. *Basic Reconstruction Algorithms for Parallel Imaging*, pages 19–36. 01 2007.

[3] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig. Coil compression for accelerated imaging with Cartesian sampling. *Magnetic Resonance in Medicine*, 69(2):571–582, 2013.

[4] Chu, A. and Noll, D.C. Coil compression in simultaneous multislice functional MRI with concentric ring slice-GRAPPA and SENSE. *Magnetic resonance in medicine*, 76(4):1196–1209, oct 2016.

[5] J. Wang, Z. Chen, Y. Wang, L. Yuan, and L. Xia. A feasibility study of geometric-decomposition coil compression in mri radial acquisitions. *Computational and Mathematical Methods in Medicine*, 2017:1–9, 01 2017.

[6] Y. Chang and H. Wang. Kernel principal component analysis of coil compression in parallel imaging. *Computational and Mathematical Methods in Medicine*, 2018:1–9, 04 2018.

[7] M. Buehrer, K. P. Pruessmann, P. Boesiger, and S. Kozerke. Array compression for MRI with large coil arrays. *Magnetic Resonance in Medicine*, 57(6):1131–1139, 2007.

[8] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[9] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

[10] M. Lustig and J. M. Pauly. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magnetic resonance in medicine*, 64(2):457–471, aug 2010.

[11] D. O. Walsh, A. F. Gmitro, and M. W Marcellin. Adaptive reconstruction of phased array MR imagery. *Magnetic Resonance in Medicine*, 43(5):682–690, 2000.

[12] S. Aja-Fernández and G. Vegas Sánchez-Ferrero. *Statistical Analysis of Noise in MRI*. 01 2016.

[13] A. Breger et al. On orthogonal projections for dimension reduction and applications in augmented target loss functions for learning problems. *accepted in Journal of Mathematical Imaging and Vision (JMIV)*, 2019.

[14] S. Dasgupta and A. Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[15] J. S. Brauchart, A. B. Reznikov, E. B. Saff, I. H. Sloan, Y. G. Wang, and R. S. Womersley. Random Point Sets on the Sphere—Hole Radii, Covering, and Separation. *Experimental Mathematics*, 27(1):62–81, 2018.

[16] A. Breger, M Ehler, and M. Gräf. Points on manifolds with asymptotically optimal covering radius. *Journal of Complexity*, 48:1–14, 2018.

[17] A. Reznikov and E. B. Saff. The Covering Radius of Randomly Distributed Points on a Manifold. *International Mathematics Research Notices*, 2016(19):6065–6094, 12 2015.

[18] Y. Chikuse. *Statistics on special manifolds*. Lecture Notes in Statistics. Springer, New York, 2003.

[19] C. A. Cocosco, V. Kollokian, R.K.-S. Kwan, G.B. Pike, and A.C. Evans. Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage*, 5:425, 1997.

[20] R. K. Kwan, A. C. Evans, and G. B. Pike. Mri simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging*, 18(11):1085–1097, Nov 1999.

[21] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, June 1998.

[22] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans. Brainweb: Online interface to a 3d mri simulated brain database. In *NeuroImage*. Citeseer, 1997.

[23] W. S. Hoge, K. Setsompop, and J.R. Polimeni. Dual-polarity slice-grappa for concurrent ghost correction and slice separation in simultaneous multi-slice epi. *Magnetic Resonance in Medicine*, 80(4):1364–1375, 2018.

[24] N. J. Tustison, Y. Yang, and M. Salerno. Advanced normalization tools for cardiac motion correction. In Oscar Camara, Tommaso Mansi, Mihaela Pop, Kawal Rhode, Maxime Sermesant, and Alistair Young, editors, *Statistical Atlases and Computational Models of the Heart - Imaging and Modelling Challenges*, pages 3–12, Cham, 2015. Springer International Publishing.

# Chapter 2

# Augmented target loss function in medical applications

The following introductory section about the loss function can be found in my paper ([1], Section 5), which was accepted in the Journal of Mathematical Imaging and Vision (JMIV), Springer (2019).

## 2.1 Augmented target loss functions [1]

Whereas the input data is processed and optimized in each iteration, the target data stays usually unchanged during the whole learning process, serving as measure of accuracy. The representation of the target data is one key property for successful approximation with neural networks. Here, we will introduce a general class of loss functions, i.e. augmented target (AT) loss functions, that use projections and features to yield beneficial representations of the target space, emphasizing important characteristics.

In optimization problems additional penalty terms are used for regularization or to enforce other beneficial constraints. In deep learning, weight decay (i.e. Tikhonov regularization) is a standard adaption of the loss function to that effect. Incorporating additional underlying information via features of the output/target data has been studied in diverse settings tailored to particular imaging applications. Perceptual loss functions have been used in [2] for image super-resolution, incorporating the comparison of high-level image features that arise from pretrained convolutional neural networks, i.e. the VGG-network [3]. Deep perceptual similarity metrics have been proposed in [4] for generating images, comparing image features instead of the original images. In [5] a similar approach was successfully used for style transfer and super-resolution, adding a network that defines loss functions. Anatomically constrained neural networks

(ACNN) have been introduced in [6] and applied to cardiac image enhancement and segmentation. Their loss functions incorporate structural information by using autoencoders to gain features about lower dimensional parametrization of the segmentation. Brain segmentation was studied in [7], where information about the desired structure has been added in the loss function via an adjacency matrix. It was used for fine-tuning the supervised learned network with unlabeled data, reducing the number of abnormalities in the segmentation.

The information of certain target characteristics can be very powerful and even replace the need of annotations in some tasks. In [8] label-free learning is approached by using just structural information of the desired output in the loss function instead of annotated target values.

In the following, we will define a general framework of loss functions that add information of target characteristics via features and projections in supervised learning tasks.

### 2.1.1 General framework

Let the training data be input vectors $\{x_i\}_{i=1}^m \subset \mathbb{R}^r$ with associated target values $\{y_i\}_{i=1}^m \subset \mathbb{R}^s$. We consider training a neural network

$$f_\theta : \mathbb{R}^r \to \mathbb{R}^s,$$

where $\theta \in \mathbb{R}^N$ corresponds to the vector of all free parameters of a fixed architecture. In each optimization step for $\theta$, the network's output $\{\hat{y}_i = f_\theta(x_i)\}_{i=1}^m \subset \mathbb{R}^s$ is compared with the targets $\{y_i\}_{i=1}^m$ via an underlying loss function $L$.

In contrast to ordinary learning problems with highly accurate target data, complicated learning tasks arising in many real world problems do not yield sufficient results when optimizing neural networks with standard loss functions $L$, such as the widely used mean least squares error

$$L_{\mathrm{MSE}}(\{y_i\}_{i=1}^m, \{\hat{y}_i\}_{i=1}^m) := \frac{1}{m} \sum_{i=1}^m \|y_i - \hat{y}_i\|^2. \tag{2.1}$$

The training data may include important information that is obvious for humans, but poorly represented within the original target data and therefore lacks consideration in the learning process. To overcome this issue, we propose to add information tailored to the particular learning problem represented by additional features of the outputs and targets.

First, we select transformations

$$T_j : \mathbb{R}^s \to \mathbb{R}^t, \quad j = 1, \ldots, d,$$

to enable error estimation in transformed output/target spaces. Note that the transformations $T_j$ are not required to be linear. However, they should be piecewise differentiable to enable subsequent optimization of the loss function with gradient methods. We shall allow for additional weighting of the transformations $T_1, \ldots, T_d$ to facilitate the selection of features for a specific learning problem.

The previous sections suggest that orthogonal projections can provide favorable feature combinations, which essentially turns into a weighting procedure.

To enable suitable projections, we stack the $d$ output/target features

$$T(y_i) := \begin{pmatrix} T_1(y_i)^\top \\ \vdots \\ T_d(y_i)^\top \end{pmatrix} \in \mathbb{R}^{d \times t},$$

so that applying a projector $p \in \mathcal{G}_{k,d}$ to each column of $T(y_i)$ yields $p(T(y_i)) \in \mathbb{R}^{d \times t}$. We now define the augmented target loss function with projections by

$$L_p\big(\{y_i\}, \{\hat{y}_i\}\big) := L(\{y_i\}, \{\hat{y}_i\}) + \alpha \cdot \tilde{L}\big(\{p(T(y_i))\}, \{p(T(\hat{y}_i))\}\big), \qquad (2.2)$$

where $\alpha > 0$ and $L$, $\tilde{L}$ correspond to conventional loss functions. Apparently, $L_p$ depends on the choice of $p \in \mathcal{G}_{k,d}$. The projection $p(T(y_i))$ weighs the previously chosen feature transformations $T(y_i)$. Standard choices of $L$ and $\tilde{L}$ are $L_{\mathrm{MSE}}$, in which case $L_p$ becomes

$$L_p\big(\{y_i\}, \{\hat{y}_i\}\big) = \frac{1}{m} \sum_{i=1}^{m} \|y_i - \hat{y}_i\|^2 + \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \|p(T(y_i)) - p(T(\hat{y}_i))\|_{\mathrm{F}}^2. \quad (2.3)$$

**Remark 2.1.1** *For $k = d$ the projector $p$ is the identity. In this case the transformations can map into different spaces, i.e.*

$$T_j : \mathbb{R}^s \to \mathbb{R}^{t_j}, \quad j = 1, \ldots, d,$$

*and we can now write the standard augmented target loss function by*

$$L_{AT}\big(\{y_i\}, \{\hat{y}_i\}\big) = \sum_{j=1}^{d} \alpha_j \cdot L^j\big(\{T_j(y_i)\}, \{T_j(\hat{y}_i)\}\big), \qquad (2.4)$$

*where $T_1$ corresponds to the identity function, $L^1, \ldots, L^d$ are common loss functions and $\alpha_1, \ldots, \alpha_d > 0$ are weighting parameters.*

It should be mentioned that $\alpha$ resembles a regularization parameter. The actual minimization of (2.1) among $\theta$ is usually performed through Tikhonov type regularization in many standard deep neural network implementations. The formulation (2.2) adds one further variational step for beneficial output data representation.

**Remark 2.1.2** *Our proposed structure with target feature maps $T_1, \ldots, T_d$ as in (2.4) relates to multi-task learning, which has been successfully used in deep neural networks [9]. It handles multiple learning problems with different outputs at the same time. In contrast to multi-task learning, we aim to solve a single problem but also penalize the error in transformed spaces enhancing certain target characteristics.*

18

## 2.2 Augmented target loss for brain tissue segmentation

The following research was developed in collaboration with Fan Zhang, PhD (Instructor, Harvard Medical School).

### 2.2.1 Introduction

Brain tissue segmentation is important in diffusion-weighted magnetic resonance imaging (dMRI) for quantification of brain microstructure and visualization tasks. The segmentation aims to label the regions of gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Tissue segmentation with T1- or T2-weighted (T1w or T2w) anatomical MRI data are based on images with high resolution and good tissue contrast. State-of-the-art dMRI segmentation is based on anatomical MRI data segmentation that is registered to the dMRI space. This requires inter-MRI modality registration, which is challenged by low image resolution [10] and echo-planar image (EPI) distortions [11].

In dMRI, specific MRI sequences are used to generate images based on the diffusion of water molecules, creating contrast in MR images. Diffusional processes are influenced by the geometrical structure of the environment and can be explored by MR non-invasively. This is especially beneficial when the size of the boundaries influencing diffusion are very small, such that it could not be resolved by conventional MRI techniques [12]. As dMRI provides important microstructural tissue information that is not available in anatomical MRI, it adds new informatin and could be useful for tissue segmentation.

Since diffusion appears to be a 3-dimensional process, molecular mobility in tissues may be anisotropic. Diffusion anisotropy effects can be fully extracted and characterized with diffusion tensor imaging (DTI). It provides further details on tissue microstructures and enables advanced application of fiber tracking in the brain [13]. Features derived from the DTI model, such as mean diffusivity (MD) and fractional anisotropy (FA), are commonly used for dMRI-based brain tissue segmentation methods, see e.g. [14], also in deep learning, see e.g. [15]. However, DTI is known to be relatively non-specific to tissue changes and can be similar between GM and WM. Diffusion kurtosis imaging (DKI) [16] is a clinically feasible extension of DTI that characterizes non-Gaussian water molecule diffusion. It enhances DTI by providing information about molecular restrictions and tissue heterogeneity in the brain.

We aim to study the application of an augmented target loss function in a deep learning method for brain tissue segmentation prediction with only dMRI data. Instead of just using DTI parameters, DKI parameters have been derived from the kurtosis tensor ([17]) and added as input features for the segmentation task. The augmented target loss penalizes misclassification of voxels during the training of a convolutional neural networks (CNN) model.

### 2.2.2 Data

For our experiments we used a high-quality HCP dMRI data set for the training, validation and testing of the neural network. Diffusion weighted images (DWIs) were acquired using a customized Connectome Siemens Skyra scanner (acquired in both AP and PA phase encoding to correct for EPI distortions), with TE/TR=89.5/5520 ms, $25 \times 1.25 \times 1.25mm^3$, and 18 b=0 images, 90 gradient directions at each of $b = 1000, 2000$ and $3000 s/mm^3$. The DWI data was processed following the HCP processing pipeline [18] and co-registered to the corresponding T1w data. Ground truth labels were obtained by using a freesurfer segmentation (computed from the T1w data) creating a label map with WM, GM and CSF. Here we use 50 subjects (30 for training, 10 for validation and 10 for testing).

For each brain we extract a 3D image with 11 features per voxel: 3 novel MK-curve features, 3 corrected DKI parameters and 5 corrected DTI parameters. These new features are now used as input for training a U-Net [19] that segments WM, GM and CSF.

### 2.2.3 Augmented target loss

We use an adapted augmented target loss function ([1]), described in Section 2.1, to train a U-Net.

The design is motivated by the fact that segmentation of CSF is relatively challenging, see e.g. [18], due to partial volume effects. This yields the loss function

$$L_{AT} = L_{CE} + \alpha \cdot L_{MSE}(x^T y, x^T t), \qquad (2.5)$$

where $L_{CE}$ corresponds to the categorical cross-entropy loss, $L_{MSE}$ is the mean-squared error loss, $\alpha$ is the weighting parameter and $x = (0.5, 0.5, 0)^T$. This corresponds to the transformation $T(\cdot) := \langle \cdot, x \rangle$ for $y$ and $t$ in the general definition (2.4).

Considering that CSF has essentially different diffusion properties (unrestricted water diffusion) from GM and WM, using the augmented target loss $L_{AT}$ adds penalization of misclassification of GM and WM as CSF and can be regulated with the parameter $\alpha$. We tested different parameters and set $\alpha = 0.5$ and $x = (0.5, 0.5, 0)^T$, where the three channels correspond to WM, GM and CSF, respectively. Including the augmented target loss improved the prediction accuracy by around 1% compared to the training of the U-Net without $L_{AT}$.

### 2.2.4 Results

In Table 2.1 we quantitatively compare the segmentation performance for different input features:

1. $F_{DTI}$, containing the 5 DTI parameters from DTI modeling (as in [24])

2. $F_{prop}$, containing the proposed 11 corrected parameters

|  | Acc | Prec | Rec |
|---|---|---|---|
| $F_{DTI}$ | 89.21% | $89.24 \pm 2.10\%$ | $88.79 \pm 3.39\%$ |
| $F_{prop}$ | 91.42% | $91.59 \pm 2.16\%$ | $91.07 \pm 1.88\%$ |

Table 2.1: Quantitative comparison between different tissue features

Comparison of the input feature sets $F_{DTI}$ and $F_{prop}$ was performed on the HCP test data set with 10 subjets, using the overall prediction accuracy (Acc), as well as mean precision (Prec) and mean recall (Rec) across the three tissue classes. The proposed U-Net with the $L_{AT}$ loss function was used for the segmentation prediction.

### 2.2.5   Discussion

The tissue segmentation learning based on our feature selection and the augmented target loss function yields promising segmentation results. We can see that the proposed DKI-based features improved GM/WM/CSF segmentation compared to using only conventional DTI features. This probably happens due to including information about restricted water diffusion properties. Moreover, we could visually observe that most of the misclassifications happens at the tissue boundaries. This yields the idea to include an augmented target loss that puts more focus on these regions in the future.

Further experiments are going on, including bigger data sets and clinical data as well as different configurations of the augmented target loss.

# Bibliography

[1] A. Breger, J. I. Orlando, P. Harar, S. Klimscha M. Dörfler, C. Grechenig, B. S. Gerendas, U. Schmidt-Erfurth, and M. Ehler. On orthogonal projections for dimension reduction and applications in augmented target loss functions for learning problems. *accepted in Journal of Mathematical Imaging and Vision (JMIV), Springer*, 2019.

[2] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017.

[3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

[4] Alexey Dosovitskiy and Thomas Brox. Generating images with perceptual similarity metrics based on deep networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 658–666, USA, 2016. Curran Associates Inc.

[5] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016.

[6] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Ricardo Guerrero, Stuart A Cook, Antonio de Marvao, Timothy Dawes, Declan O'Regan, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Anatomically constrained neural networks (acnn): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, PP, 05 2017.

[7] Pierre-Antoine Ganaye, Michaël Sdika, and Hugues Benoit-Cattin. *Semi-supervised Learning for Segmentation Under Semantic Constraint: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III*, pages 595–602. 09 2018.

[8] Russell Stewart and Stefano Ermon. Label-free supervision of neural networks with physics and domain knowledge. In *AAAI*, 2017.

[9] R. Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, Jul 1997.

[10] Milos Malinsky, Roman Peter, Erlend Hodneland, Astri Lundervold, Arvid Lundervold, and Jiří Jan. Registration of fa and t1-weighted mri data of healthy human brain based on template matching and normalized cross-correlation. *Journal of digital imaging*, 26(4):774–785, 08 2013.

[11] Angela Albi, Antonio Meola, Fan Zhang, Pegah Kahali, Laura Rigolo, Chantal M.W. Tax, Pelin Aksit Ciris, Walid I. Essayed, Prashin Unadkat, Isaiah Norton, Yogesh Rathi, Olutayo Olubiyi, Alexandra J. Golby, and Lauren J. O'Donnell. Image registration to compensate for epi distortion in patients with brain tumors: An evaluation of tract-specific effects. *Journal of Neuroimaging*, 28(2):173–182, 2018.

[12] Heidi Johansen-Berg and Timothy E.J. Behrens, editors. *Diffusion MRI (Second Edition)*. Academic Press, second edition edition, 2014.

[13] Denis Le Bihan, Jean-François Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546, 2001.

[14] Tianming Liu, Hai Li, Kelvin Wong, Ashley Tarokh, Lei Guo, and Stephen T C Wong. Brain tissue segmentation based on dti data. *NeuroImage*, 38(1):114–123, 10 2007.

[15] Wenlu Zhang, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, and Dinggang Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage*, 108:214–224, 03 2015.

[16] Lauren J O'Donnell and Ofer Pasternak. Does diffusion mri tell us anything about the white matter? an overview of methods and pitfalls. *Schizophrenia research*, 161(1):133–141, 01 2015.

[17] Andrew J. Steven, Jiachen Zhuo, and Elias R. Melhem. Diffusion kurtosis imaging: An emerging technique for evaluating the microstructural environment of the brain. *American Journal of Roentgenology*, 202(1):W26–W33, 2019/08/15 2013.

[18] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, David C Van Essen, Mark Jenkinson, and WU-Minn HCP Consortium. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 10 2013.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

## 2.3 Augmented target loss for photoreceptor layer segmentation in OCT scans

# An amplified-target loss approach for photoreceptor layer segmentation in pathological OCT scans

Anna Breger[1][*], José Ignacio Orlando[2][*], Hrvoje Bogunović[2], Sophie Riedl[2], Bianca S. Gerendas[2], Martin Ehler[1], and Ursula Schmidt-Erfurth[2]

[1] Christian Doppler Laboratory for Ophthalmic Image Analysis (OPTIMA), Department of Ophthalmology and Optometry, Medical University of Vienna, Austria
[2] Department of Mathematics, University of Vienna, Austria

**Abstract.** Segmenting anatomical structures such as the photoreceptor layer in retinal optical coherence tomography (OCT) scans is challenging in pathological scenarios. Supervised deep learning models trained with standard loss functions are usually able to characterize only the most common disease appearance from a training set, resulting in suboptimal performance and poor generalization when dealing with unseen lesions. In this paper we propose to overcome this limitation by means of an augmented target loss function framework. We introduce a novel amplified-target loss that explicitly penalizes errors within the central area of the input images, based on the observation that most of the challenging disease appearance is usually located in this area. We experimentally validated our approach using a data set with OCT scans of patients with macular diseases. We observe increased performance compared to the models that use only the standard losses. Our proposed loss function strongly supports the segmentation model to better distinguish photoreceptors in highly pathological scenarios.

## 1   Introduction

Supervised deep learning techniques have revolutionized the field of medical image segmentation [1], particularlly with fully convolutional neural network architectures such as the U-Net [2]. To learn these networks, a loss function $L$ is optimized using gradient based approaches and backpropagation. This function is usually defined in terms of metrics that quantify the discrepancies between a trustworthy/ground truth labelling and the predicted segmentation.

In this typical framework a loss function is not explicitly tailored to aim for a specific feature in the target space. Hence, the network firstly learns the dominating characteristics of the target images in the training set, and its remaining capacity is gradually devoted to characterize other less prevalent target features. This becomes an issue when dealing with highly pathological data, where lesions or disease appeareance might significantly differ between patients. To overcome

---

[*] equal contribution

this limitation, some authors proposed to train segmentation models using a linear combination of different losses such as cross-entropy and Dice [3]. However, these metrics are still computed from the same target representation, so they do not enhance a specific target feature. In this paper we propose to extend this idea by using the framework of *augmented target loss functions*, introduced in [4]. Rather than relying on a single or a linear combination of loss functions defined on the original prediction and target space, Breger *et al.* [4] proposed to compute the loss on alternative representations of the predictions and targets, obtained by applying differentiable transformations $T$ that enhance specific characteristics.

This paper focuses on the application of an augmented target loss function for photoreceptor layer segmentation in retinal optical coherence tomography (OCT) scans of patients with macular diseases. OCT is the state-of-the-art technique for imaging the retina, as it brings volumetric information through a stack of 2D scans (B-scans) at a micrometric resolution [5]. Ophthalmic disorders such as diabetic macular edema (DME), retinal vein occlusion (RVO) and age-related macular degeneration (AMD) gradually affect photoreceptors while progressing. The abnormal accumulation of fluid due to these diseases significantly alters the retina, eventually leading to photoreceptor cell death. This last characteristic can be noticed through OCT imaging: first as a pathological thinning of the photoreceptor layer, and more lately as complete disruptions on it (Fig. 1, right). It has been observed that these abnormalities are highly correlated with focal vision impairment [6] and visual accuity loss when located at the central area of the retina [7]. Hence, the automated characterization of the morphology of the photoreceptor layer is relevant for efficient quantification of functional loss.

In this paper we build on top of the architectural innovations proposed in [8] by training such a model using an augmented target loss function. Fitting the framework we introduce a novel amplified-target loss that induces further penalization to errors within the central area of the B-scans. As the most challenging pathologies are usually observed at the central area of fovea-centered OCT scans, our hypothesis is that incorporating this loss function as a kind of regularizer enforces the network to better characterize disease appeareance. We validate our approach using a series of OCT scans of patients with AMD, DME and RVO. Our results empirically show that the proposed loss functions improve the performance within the central millimeters of the retina compared to using traditional losses without compromising the performance in the entire volume.

## 2   Methods

### 2.1   Augmented target loss functions for image segmentation

In a supervised learning problem we aim to learn a function $f$ with $f_\theta(x) \approx y$, where $\theta$ denotes the free parameters and $S = \{(x, y)^{(i)}\}, 1 < i < N$ is a given training set with pairs of inputs $x$ and ground truth labels $y$. In the context of image segmentation, $x$ corresponds to an input image, $y$ and $\hat{y}$ are manual and predicted segmentations and $f_\theta$ is some segmentation model (e.g. a fully convolutional neural network such as the U-Net [2]).
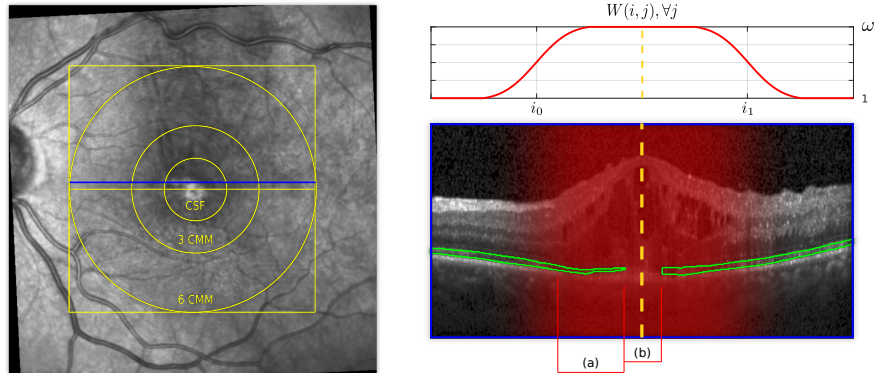
**Fig. 1.** Left: scanning laser ophthalmoscopy (SLO) of a patient with RVO. The square indicates the area captured by the OCT volume and the rings represent the central subfield (CSF) and the 3 and 6 central millimeters (3 CMM and 6 CMM). The blue line highlights the B-scan showed in the right side. Right: CSF B-scan with photoreceptor layer annotation (green) with (a) disruptions and (b) abnormal thinning. The red heat map represents the weighting strategy applied in our loss function. The central coordinate of the image is indicated with the yellow dotted line, and a profile of the weighting strategy is illustrated on top of the B-scan.

To adjust the weights $\theta$ from the chosen network structure $f_\theta$, a loss function $L$ is minimized using gradient based optimization. $L$ is a piecewise differentiable loss function, e.g. cross-entropy (CE) or mean square error (MSE), that measures the pixel-wise differences between $\hat{y}$ and $y$. In standard settings no specific areas of the images are penalized more than others. Thus, the parameters $\theta$ are mostly adjusted to characterize those features from the training set that have the most impact on the overall loss. Although this might be helpful to segment healthy anatomy, in pathological scenarios the network will overfit the prevalent features unless explicit regularization is imposed during training.

Here, we propose to use the framework of augmented target (AT) loss functions, introduced in [4]. These losses take into account prior knowledge of target characteristics via error estimation in transformed target spaces. The framework can be applied to any supervised learning problem based on loss optimization where additional information about the target data is available, provided it can be formulated as a transformation function $T$. The transformation may correspond to any piecewise differentiable function on the target space that yields a more beneficial representation of some known target characteristic.

Following [4], the AT loss functions $L_{\mathrm{AT}}$ is a linear combination of losses applied to transformed targets. Its general form is:

$$L_{\mathrm{AT}} = \sum_{j=1}^{d} \lambda_j \cdot L^j \big( \{T_j(y_i)\}, \{T_j(\hat{y}_i)\} \big), \tag{1}$$

where $\lambda_j > 0$ corresponds to some weight, $T_j$ to a specific transformation and $L^j$ to some loss function, for all $j \in \{1, \ldots, d\}$.

Setting typically $T_1$ to the identity and $L^1$ to a standard loss, the additional terms in the $L_{\text{AT}}$ loss act as as amplified target information, yielding a new optimization problem:

$$\hat{\theta} = \arg\min_\theta \{\lambda_1 \cdot L_1\big(\{y_i\}, \{\hat{y}_i\}\big) + \sum_{j=2}^d \lambda_j \cdot L^j\big(\{T_j(y_i)\}, \{T_j(\hat{y}_i)\}\big), \qquad (2)$$

where the weights $\lambda_1$ and $\{\lambda_j\}_{j=2}^d$ control the balance between the main loss and the regularization terms respectively.

## 2.2 Amplified-target loss functions for photoreceptor layer segmentation

We experimentally study the AT loss function framework in the context of photoreceptor layer segmentation in pathological OCT scans. We tailor a so called *amplified-target loss* in which a transformation $T$ is designed to bring an increased penalty to errors within the central area of the images. This loss is intended to incorporate the prior knowledge that abnormalities such as pathological thinnings and disruptions of the photoreceptor layer are more common in the central millimeters of the foveal area. To do so, we define a transformation $T(y_i) = \langle y_i, W \rangle$, where $y_i$ corresponds to the given binary targets and $W$ represents a weighting matrix that encodes a penalization weight for errors. This operation can analogously be applied to the predictions $\hat{y}_i$. Fig. 1 graphically illustrates the design of the weighting matrix $W$. Formally, we define $W = G_\sigma * V$, where $G_\sigma$ stands for a Gaussian filter with standard deviation $\sigma$. We define $V$ as:

$$V_{i,j} := \begin{cases} \omega & \text{for } i_0 < i < i_1 \text{ and all } j, \\ 1 & \text{otherwise,} \end{cases} \qquad (3)$$

where $\omega$ denotes the maximum weight assigned to the central area and $[i_0, i_1]$ is the horizontal interval of the image that is amplified. The Gaussian filter $G_\sigma$ is used to smooth the penalization factor within the edges of the interval.

Following the formulation in (2), we can then redefine our empirical risk minimization problem as

$$\hat{\theta} = \arg\min_\theta \{\lambda_1 \cdot L^1\big(\{y_i\}, \{\hat{y}_i\}\big) + \lambda_2 \cdot L^2\big(\{\langle y_i, W \rangle, \langle \hat{y}_i, W \rangle\}\big)\}, \qquad (4)$$

where we choose $\lambda_1, \lambda_2 \in \mathbb{R}$ and $L^1 = L^2$ as CE or MSE losses.

## 3 Experimental setup

### 3.1 Materials

Our method was trained and tested on an in-house data set with 53 Spectralis OCT volumes of patients suffering from DME (10), RVO (27) and AMD (10).

Each image comprises $496 \times 512$ pixels per B-scan, 49 B-scans per volume. All the B-scans were manually annotated by certified readers under the supervision of a retina expert, who modified the labels when necessary to obtain ground truth segmentations. The set was randomly divided into a training, a validation and a test set, each of them with 34, 4 and 15 scans, respectively, with approximately the same distribution of diseases and percentages of disrupted columns per B-scan (or A-scans).

### 3.2   Network architecture and training setup

We used the photoreceptor segmentation network described in [8] in our experiments (note that any other architecture could be applied within our framework). We used as baselines CE and MSE comparing it to the adapted AT loss.

Every configuration was trained at a B-scan level with a batch size of 2 images, using Adam optimization and early stopping. Hence, training was stopped if the validation loss did not improve for the last 45 epochs. The learning rate was set to $\eta = 0.0001$, and divided by 2 if the validation loss was not improved during the last 15 epochs. Data augmentation was used in the form of random horizontal flippings. Binary segmentations were retrieved as in [8] by thresholding the softmax scores of the photoreceptors class using the Otsu algorithm.

## 4    Results and Discussion

We evaluated the performance for segmenting the photoreceptor layer using the volume-wise Dice index, at the CSF, the 3 CMM, the 3-1 ring and the full volume (Fig. 1, left). All the experiments with our AT loss functions were conducted using fixed values for $\sigma = \frac{1}{16}X$, $i_0 = \frac{1}{4}X$ and $i_1 = \frac{3}{4}X$ (with $X = 512$ being the horizontal size of the B-scans, in pixels), without optimizing them on the validation set. Different configurations for $\omega = 2^k, k \in \{1, ..., 5\}$ and $\lambda_1$ and $\lambda_2 \in \{0.001, 0.01, 0.1, 1, 2, 4, 8\}$ were analyzed, and the best configuration according to Dice index on the validation set was then fixed to allow a fair comparison on the test set. From this model selection step, we observed that $\omega = 8$, $\lambda_1 = 1$ and $\lambda_2 = 8$ reported the best performance for the AT loss with categorical cross-entropy (CE), and $\omega = 32$, $\lambda_1 = \lambda_2 = 1$ for the AT loss with mean square error (MSE).

Fig. 2 depicts boxplots with the quantitative performance of each model on the test set, compared with their corresponding baselines trained only with CE and MSE, for each evaluation area. The mean and standard deviation values of the Dice index are presented in Table 1. The incorporation of the AT loss allows to perform consistently better in all the cases, with the best results reported by the MSE loss. Statistical analysis using one-tail Wilcoxon sign-rank tests at a significance level $\alpha = 0.05$ showed that the model trained with MSE + AT loss reported significantly higher Dice values in the CSF area compared to using CE + AT loss or only MSE ($p < 0.0171$). These differences were not statistical significant with respect to the model trained with CE ($p = 0.1902$).
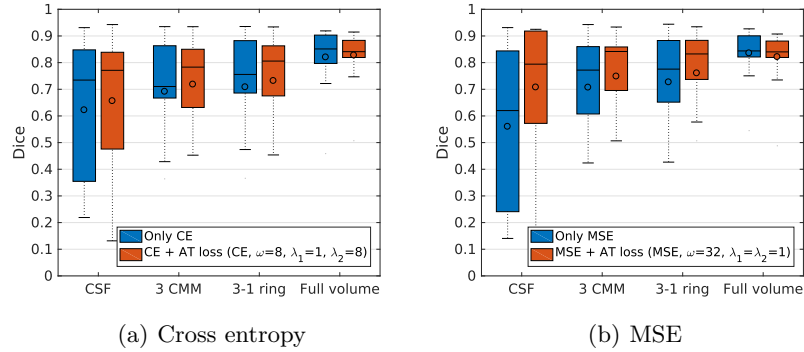
(a) Cross entropy                    (b) MSE

**Fig. 2.** Volume-wise Dice values for all the evaluated models and our proposed approach in each evaluation area. Circles indicate mean values. CSF: central subfield (1 central millimeter). 3 CMM: three central millimeter. 3-1 ring: area between CSF and 3 CMM.

**Table 1.** Volume-wise mean $\pm$ standard deviation Dice values in the test set for each photoreceptor segmentation model in each area.

| Method | CSF | 3 CMM | 3-1 ring | Full volume |
|---|---|---|---|---|
| CE loss | 0.622 $\pm$ 0.271 | 0.691 $\pm$ 0.242 | 0.708 $\pm$ 0.242 | 0.820 $\pm$ 0.118 |
| CE + AT loss (CE, $\omega = 8$, $\lambda_1 = 1$, $\lambda_2 = 8$) | **0.656** **$\pm$ 0.256** | **0.718** **$\pm$ 0.218** | **0.732** **$\pm$ 0.218** | **0.828** **$\pm$ 0.100** |
| MSE loss | 0.560 $\pm$ 0.303 | 0.707 $\pm$ 0.223 | 0.727 $\pm$ 0.223 | **0.835** **$\pm$ 0.096** |
| MSE + AT loss (MSE, $\omega = 32$, $\lambda_1 = \lambda_2 = 1$) | **0.708** **$\pm$ 0.254** | **0.749** **$\pm$ 0.215** | **0.760** **$\pm$ 0.213** | 0.821 $\pm$ 0.102 |

When comparing the Dice values at the 3-1 ring, the MSE with AT loss model reported statistically significant better results than using only CE or MSE ($p < 0.0042$), which is consistent with its behavior in the 3 CMM ($p < 0.0416$). No statistically significant differences in performance were observed at the full volume level (two-tails test, $p > 0.0730$).

We qualitatively analyzed the segmentation and score maps using the CE and MSE combined with the AT loss. Fig. 3 depicts segmentation results in a central B-scan from the test set, with score maps represented as heatmaps. Using MSE produces noisy scores within the lateral areas of the B-scans, and therefore spurious elements in the segmentation. CE, on the contrary, results in smoother score maps, although with few false negatives in the vecinity of subretinal fluid. This behavior is linked to the one observed in Table 1, where the MSE + AT loss model reported higher Dice in the central area than using CE, and smaller values in the full volume. The model trained with only MSE performs poorly in the CSF, the 3 CMM and the 3-1 ring, which indicate that it struggles to deal with pathologies. Similarly, the high performance at a volume level indicates that it
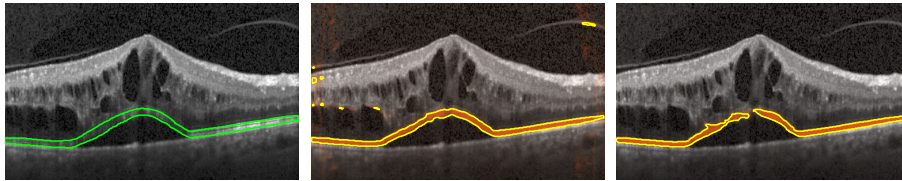
**Fig. 3.** Qualitative effect of the loss selection in the pixel score values. From left to right: manual annotation (green), score map (orange) and binary segmentation (yellow) obtained with MSE + AT loss (MSE, $\omega = 32, \lambda_1 = \lambda_2 = 1$) and CE + AT loss (CE, $\omega = 8$, $\lambda_1 = 1$, $\lambda_2 = 8$).

can better characterize normal appeareances. When using MSE + AT loss, a significant reduction in the amount of false negatives occurs at the central areas. However, as mentioned before, the score maps are noisy at the borders of the B-scans, which causes a drop in the full volume Dice. The model trained with CE + AT loss is less accurate at the center than the one trained with MSE + AT loss, but it still outperforms the baseline approaches. Moreover, at a volume basis the CE + AT loss remains competitive with respect to the one trained only with CE loss.

Finally, Fig. 4 presents qualitative results in exemplary central B-scans from our test set obtained both by the models trained with CE only and with CE + AT loss. Our approach produced more anatomically plausible segmentations than the standard CE loss in pathological areas with subretinal fluid (Fig. 4 (a) and (b)) or large disruptions (Fig. 4 (c)).

## 5    Conclusions

In this paper we proposed to use the framework of augmented target loss functions for photoreceptor layer segmentation in pathological OCT scans. We define an amplified-target loss incorparating a transformation that weights the central area of the input B-scans to further penalize errors commited in this region. We experimentally observed that this straightforward approach allows to significantly improve performance within the central millimeters of fovea-centered OCT scans, without affecting the overall performance in the entire volume. These results indicate that the proposed AT loss function acts as a form of regularization, better characterizing photoreceptors appeareance within highly pathological regions. We are currently exploring new alternatives to identify the regions to weight and to learn their corresponding weights. Further experiments are also performed to evaluate our approach in the context of other OCT based applications such as fluid segmentation and using OCT scans from other vendors.
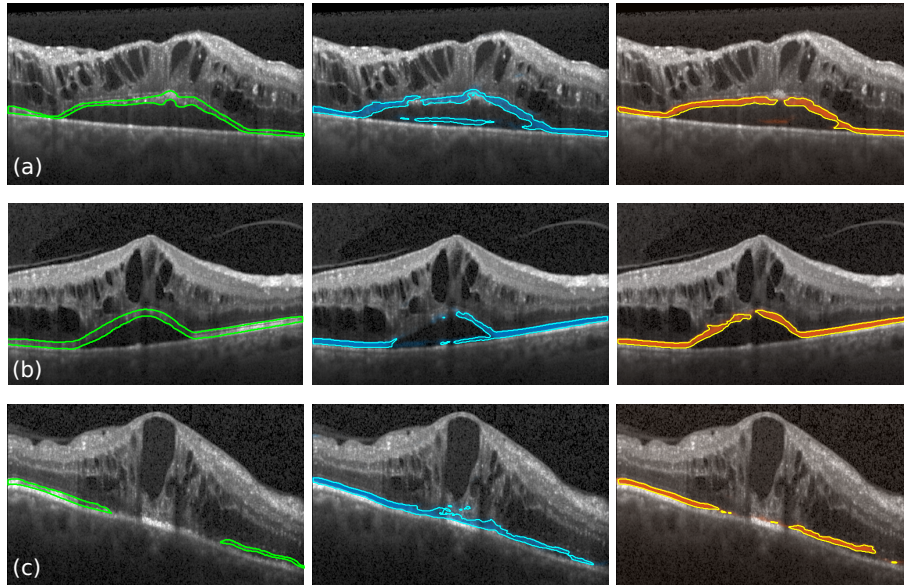
**Fig. 4.** Qualitative results in central B-scans from the test set. From left to right: manual annotations (green), results with only CE loss (blue) and results with CE + AT loss (CE, $\omega = 8$, $\lambda_1 = 1$, $\lambda_2 = 8$).

# References

1. Litjens, G., et al.: A survey on deep learning in medical image analysis. Medical image analysis **42** (2017) 60–88
2. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., eds.: MICCAI 2015, Springer (2015) 234–241
3. Khened, M., Kollerathu, V.A., Krishnamurthi, G.: Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical image analysis **51** (2019) 21–45
4. Breger, A., et al.: On orthogonal projections for dimension reduction and applications in augmented target loss functions for learning problems. accepted in Journal of Mathematics in Imaging and Vision (JMIV), Springer (2019)
5. Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B.S., Waldstein, S.M., Bogunović, H.: Artificial intelligence in retina. Progress in Retinal and Eye Research (aug 2018)
6. Takahashi, A., et al.: Photoreceptor damage and reduction of retinal sensitivity surrounding geographic atrophy in age-related macular degeneration. American journal of ophthalmology **168** (2016) 260–268
7. Gerendas, B.S., et al.: OCT biomarkers predictive for visual acuity in patients with diabetic macular edema. IOVS **58**(8) (2017) 2026–2026
8. Orlando, J.I., et al.: U2-Net: A Bayesian U-Net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological OCT scans. In: ISBI. (2019)