Internship in XLIM Laboratory, under the title:

Artificial intelligence for enhanced non-invasive medical imaging

Submitted by Marwan AL OMARI

Supervised by

Pascal BOURDON

Master of Conferences (lecturer), University of Poitiers

Master Connected Objects

SP2MI, UFR Fundamental and Applied Sciences

University of Poitiers

Internship Period: 28 Mars, 2022 - 01 September 2022

Academic year

2021 - 2022

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

In this section, abbreviated and acronym terms, which used in this paper, are clarified and presented as follows:

Machine learning (ML)

Deep learning (DL)

Normalized Mean Absolute Error (NMAE)

Peak signal-to-noise ratio (PSNR)

Structural similarity index measurement (SSIM)

Visual information fidelity (VIF)

Naturalness image quality evaluator (NIQE)

Brain Tumor Segmentation Challenge (BRATS)

The CelebFaces Attributes (CelebA)

The Radboud Faces Database (RaFD)

T1-weighted (T1)

T1 and contrast-enhance (T1c)

T2-weighted (T2)

Fluid-attenuated inversion recovery (Flair)

Convolutional neural networks (CNNs)

Magnetic resonance images (MRI)

Generative adversarial networks (GANs)

XLIM Laboratory and Internship Presentations

In this section, a brief introduction is provided on XLIM laboratory as well as its specialties. Then, an overall presentation of the internship position will be discussed, its various tasks, and managemental part.

1.1 XLIM Laboratory

XLIM is one of the French research centers that they are known under the name of CNRS (*Centre national de la recherche scientifique*/Scientific Research National Center) [7]. It is numbered as 7252. The laboratory is greatly centered on electronics, optics, photonics, mathematics, computer science, image processing, telecommunication, network security, bioengineering, and energy.

1.2 Internship Presentations

In my internship training, I am an essential research engineer to develop and experiment DL architectures for solving the problem of generating different modalities out of solely modality on Magnetic Resonance Imaging (MRI) open-source dataset.

To begin with, I am responsible to analyze the dataset in manipulations, using python programming language in *Jupyter Notebook*. Afterward, I have to develop GANs. Therefore, I have had to identify the state-of-the-art solutions and algorithms for enhancing the results, including pre-processing methods and algorithms structures.

During fact-to-face meetings, I am presenting the results that I came up with in each step with clear presentation.

1.2.1 Project Management

In this section, I am going to present the software I am using in the internship.

1.2.1.1 Operating systems

Linux and **Windows** operating systems (OS) are system software that manages computer hardware, software resources, and provides common services for computer programs. So, it is considered the fundamental software that could manage the other software that we use, for example, anaconda.

1.2.1.2 Software programs

SourceSup service, which operated through RENATER, is a management web platform for Higher Education and French Research organizations. Every member of the network can create a project at the platform in additionally permitting the collaboration of outside people at the projects.

GanttProject break down the work, build a Gantt chart, assign resources, and calculate project costs. In our case, we used to debrief the tasks for the ongoing weeks through performed and still in progress tasks. As in the following figure, we could see the current, finished, and continuous tasks that I am working on. For example, from the period between 28/03/2022 to 22/04/2022, I am covering the state-of-the-art concerning GANs and medical synthesis.

Figure 1- GANT diagram - Time management of tasks
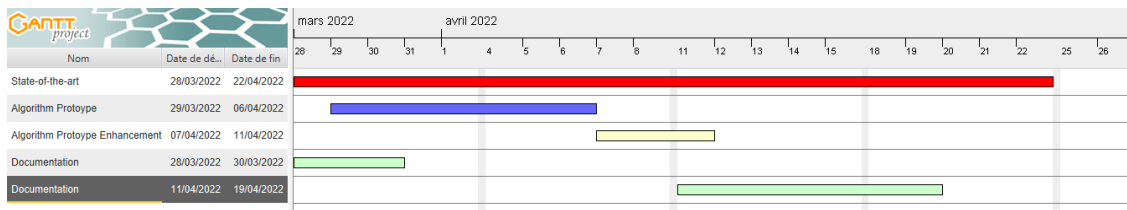
Draw.io is free online diagram software that is used to write and simplify the coding and information into a visual schema that is easy to be understood.

Webex, by Cisco, is a software for video conferencing, online meetings, screen share, and webinars. I used it to report results, problems, and complains to the supervisor as well as for planning the weekly meetings.

ENT Zimbra is a collaborative messaging to write collaborators of internship and others, share the messages, create appointments.

Microsoft Office is a text processing application that allows the configurating the attributes of a document, such as layout, the styles of content, and to add their content in a variety of ways and formats to produce documents. It is used to write the weekly reports as well as the final report for the whole internship.

Microsoft PowerPoint is a presentation programs uses slides to convey information rich in multimedia.

Zotero helps to collect, organize, cite, and share your research sources.

1.2.1.3 Coding Plateforms
Juypter Notebook is a project that provides services for interactive computing using python. Therefore, as coding in python, Juypter will visualize the results on the same script window.

Anaconda is a distribution of Python programming language for scientific computing including data science, machine learning applications, large-scale data processing, predictive analytics, etc.… It is used as for the aim to simplify package management and deployment.

1. Introduction

Image processing employs different and multiple methodologies from hand-forged features to ML for the objectives of detection, classification, and segmentation. DL employs deeper processing units called neurons to learn hierarchical features and representations of data [8]. DL outperforms humans' predictions as it is well going into impressive advancement and evolvement in the different fields of science as in computer vision [9].

In the context of medical imaging, it is the same case with fine-tuning in the proposition and auto generative tasks. As for the aim of this internship, we have, therefore, adopted GANs model to generate different attributes from a single attribute as an input. The data is in widely used for the improvements of brain lesions diagnosis. It includes patient file images; each contains the four modalities (attributes):

2. Related Work

In this section, we would present the main research articles that experimented on open-source datasets from different challenges, for example, CelebA, RaFD or BRATS2015, using GANs for the aim of better understanding the literature and the employed techniques in their approaches with respect to our project study.

[3] Their multimodal GANs network assessed on four contrasts: T1, T1c, T2, and Flair by calculating NMAE, PSNR, SSIM, VIF, NIQE. They adopted StarGAN strategy in training a model to learn all the mappings among the four MRI modalities through SISO-type method in which training 12 models to map one-to-one modality, and a sole model, trained using the unified GAN, to map all of them together. Their method was evaluated on the BRATS2015, which consists of 274 subjects (in which, 54 patients with low grade glioma, and 220 patients with high grade glioma) with four spatially modalities including T1, T1c, T2, and Flair. The voxel size of each image is $240 \times 240 \times 155$. For each image, patches with a size of $72 \times 72 \times 72$ were first extracted and fed into the model. Then, the final estimation of overlapped regions was set to 48x48x48 with 5-fold cross validation. One of the important parameters are set to $\alpha=2$, $\beta=5$, $\gamma=10$, $\delta=2$, and $\mu= 0.1$ for balancing the weights of classification loss, synthetic consistency loss, cycle consistency loss, adversarial and classification losses, and cycle consistency loss, respectively. As a conclusion, for example, with T1 as input modality, the NMAEs for the generated T1c, T2, Flair respectively are 0.034 ± 0.005, 0.041 ± 0.006, and 0.041 ± 0.006, the PSNRs are 32.353 ± 2.525 dB, 30.016 ± 2.577 dB, and 29.091 ± 2.795 dB, the SSIMs are 0.974 ± 0.059, 0.969 ± 0.059, and 0.959 ± 0.059, the VIF are 0.750 ± 0.087, 0.706 ± 0.097, and 0.654 ± 0.062, and NIQE are 1.396 ± 0.401, 1.511 ± 0.460, and 1.259 ± 0.358, respectively. Their study is limited in spatially co-registered multimodal images before even they were used for training and testing on the small dataset. For future work, they will try to increase the amount of training and testing images through augmentation techniques.

[2] StarGAN, a unified GAN, handles multi-domain image-to-image translations among multiple domains in using a single model of one-to-one discriminator generator. A single model takes in a training data of multiple domains as input both an image and domain information by one-hot encoding, and learns the mappings between all available domains, using only one generator. Training processing generates a target domain label randomly as to finish image translation. Due to the limitations in different models in treating image translation as a mono task that could not be generalized to more than an individual task, the researcher proposed a novel approach that could handle multiple domains using single generator and discriminator.

Their approach works fine on facial feature transfer and facial expression synthesis tasks. Their architecture was tested on two datasets. Firstly, CelebA dataset contains 40 labels related to facial features such as hair color, gender, and age. Secondly, RaFD dataset contains 8 labels for facial expressions as like happy, angry, and sad. Besides, to avoid missing values in the dataset, they applied a mask vector of domain label that works on ignoring unknown labels and focus on what is already available. On CelebA, StarGAN achieved 66.2%, 39.1%, 70.6%, 47.4%, 61.5%, 49.8%, 52.2% for hair color(H), gender(G), aged(A), H+G, H+A, G+A, and H+G+A, respecivly. On RaFD: StarGAN achieved 2.12 loss with $53.2M \times 1$ parameters. On both CelebA+RaFD, their model properly learned the intended role of a mask vector in image-to-image translations when involving all the labels from multiple datasets altogether. In conclusion, StarGAN generated images of higher visual quality compared to existing methods.

[13] Pix2pix differs from other architectures by including U-Net for the generator and convolutional PatchedGANs (CGANs) for their discriminator. As a supervised learning problem, pix2pix combines an adversarial loss with a L1 loss for capturing the low frequencies. Their model progressively downsamples the input images from a high-resolution grid to a bottle neck layer of skip connectors (concatenates the channels from previous layers) followed by U-Net. Their algorithm proved to be effective as the PatchGAN runs patches that could capture high frequencies on image structure. For algorithm's optimization, they had used minibatch SGD and Adam solver, with a learning rate of 0.0002, and momentum parameters $\beta 1 = 0.5$, $\beta 2 = 0.999$. Also, batch normalization had been used depending on the experiment, ranging from 0 to 10. They trained their CGANs on different tasks and datasets, including Cityscapes datasets, GMP Facades, Google Maps photos, BW to color photos, edges of photos, human-drawn sketches, day to night images, thermal to color dataset, and finally missing pixels photos top inpainted photo. They declared even though with small training size; they were able to achieve a descent result on a single Pascal Titan X GPU. As a limitation to measure structure losses in applying "traditional metroces" as per-pixel mean-squared error, they applied joint method. First, a test of map generation, aerial phot generation, and image colorization conducted for solving graphic problems on Amazon Mechanical Turk (AMT). Secondly, a metric system used for recognizing the realistically match of objects in images. As a conclusion, L1 loss, CGAN, and L1+CGAN achieved on cityscapes 86%, 76%, and 0.83 per-pixel accuracy, and 42%, 28%, and 36% per-class accuracy, respectively.

[14] UNIT combines variational autoencoders (VAEs) with CoCAN, where two generators share weights to learn the shared distribution of images in cross domain. Shred-latent space implies cycle-consistency consistency (CC) mapping between source and target domains as the translated image in target domain can be mapped back to the original domain in the target domain. They employed weight-sharing constraint to relate VAEs by the last weights of the last few layers responsible for extracting encoding0 high frequencies of images and share the for decoding these frequencies. They used ADAM with learning rate of 0.0001 and momentums set to 0.5 and 0.999. The batch consists of one image from the domain and another from the other.

[15] CycleGAN and [16] DiscoGAN saves information attributes among input and translated images by using a cycle consistency loss. However, they all suffered from limited generalization as it is SISO models.

3. Background
In this section, we would discuss the DL in the context of medical diagnosis in highlighting the GANs architectures as well as detailed description of the dataset.

3.1 Deep Learning in Medical Diagnosis
Medical image synthesis is an alternative to multiple pulse sequences for acquiring multiple contracts MRI [3]. Mainly, MRI is used in clinical practice due to its capacity in providing useful information, for example the four contracts:

- T1-weighted (T1) distinguish white and grey matters.
- T1-weighted and Contrast-enhance (T1c) assess the change of tumor shape with enhanced demarcation around tumor.
- T2-weighted (T2) shows fluid obviously from cortical tissue.
- fluid-attenuated inversion recovery (Flair) shows contours of lesion.

Furthermore, MRI scan is non-standardized process across the different institution, cross-modal image synthesis proposed to tackle this challenge in providing missing modalities.

1. Single-input single-output (SISO): a target image generated from a given source image.
2. Multi-input single-output (MISO): overcome limitations of SISO when source and target images are weakly correlated through learning shared latent representations.
3. Multi-input multi-output (MIMO): synthesizing one or more modalities from an input of MRI modalities.
4. Single-input multi-output (SIMO): where only single modality is available as input, but multiple contrasts are necessary in output.

Figure 2- SISO and SIMO methods in handling MRI modalities
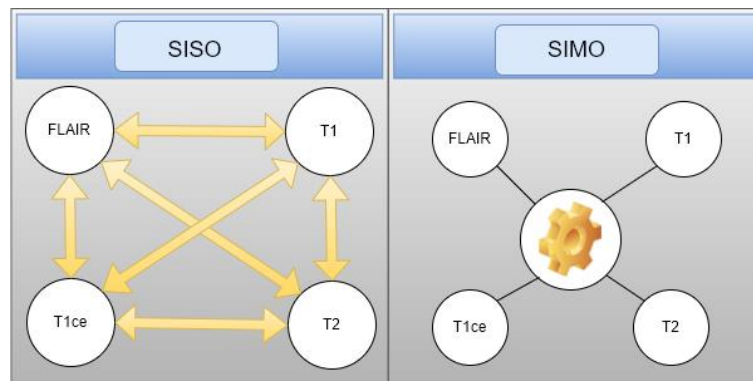
MRI of different modalities could provide complementary information for medical diagnosis, but also it is challenging and costly expensive to access all the modalities. Many methods focus on modality to modality synthesize, which marks a grand limitation in generalizing the outcomes to other modalities. Therefore, for each two modalities, we should develop a separate model so we could be able to map them together. To address this challenging problem, we propose a multiple modalities GANs to synthesize three MR modalities (FLAIR, T1 and T1ce) from one MR modality T2.

3.1.1 Generative Adversarial Networks (GANs)
By specifying a high-level target, in producing indistinguishable results from reality, and then automatically learning an appropriate loss function to satisfy that goal, is exactly what GANs

does. The loss learning GANs tries to classify whether the output image is real or fake by training a generalized model to minimize this loss. Therefore, blurred images will not be accepted as they will obviously look fake. Because GANs learn an adaptive loss to data, they can be applied to a multitude of tasks that traditionally require different types of loss functions.

As a definition, GANs are generative modeling approach using DL methods as in CNNs. In practice, GANs train a generative model by two sub-models: the generator model at one hand, learn to generate new examples by mapping the modalities of the networks' inputs as on the other hand to feed it to discriminator model in classifying the examples as either real or fake.

As we could see in the following figure, GAN generator model is considered as unsupervised learning problem as it generates a batch of samples and theses entities with the real examples are fed into the GAN model. Then, the discriminator model, as considered supervised learning, got updated based on its performance of classification of the samples to either real or fake.

Figure 3- GANs concept
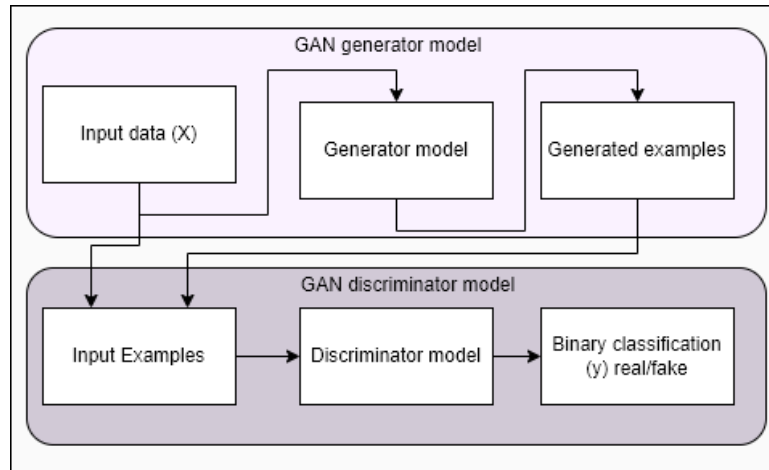
A unified GANs of single generator and discriminator to map images from four modalities. Generator takes an image with its modality as input to synthesize to target modality. The discriminator is capable to differentiate between real and synthesized images on their corresponding modalities.

GANs learn a mapping from the observed image x and the random noise vector z, toy, G: {x, z} → y. Generator G is trained to produce outputs that are indistinguishable from "real" images by a highly trained discriminator, D, trained to do the best they can to generate show "fake" generator. This training procedure is mathematized in Figure 2.

On the other hand, U-NET is used to image segmentation as it helps to reduce the volume of data. The U-Net is composed of two channels. The first channel resembles to the encoder as it captures the context of the image. On the other hand, the decoder is a transposed convolution that works to construct the original image based on image segments extracted in the process. Basically, the network assembles layers of convolution followed by max pooling later that reduce image density as to reduce the training parameters of network.

3.2 Dataset
[4][5][6] The dataset is the Brain Tumor Segmentation Challenge (BRATS2018)

4. Preliminary Experiments and Results

Training the GANs on 55 data examples in specifying these data splitting as 10 test, 22 train, 23 GAN samples, we have obtained the following results, for first experiment, at the 124 Epoch.

Table 1- Results of GANs on a small test sample of BRATS2018 dataset

d_loss	g_loss	s_loss	d_real	d_fake	d_cls	d_gp	g_fake	g_cls	g_rec	g_seg
-2.23	243.7	0.02	-4.7	2.21	0.1	0.01	-2.2	0.68	243.2	0.023

In figure (4), we could use a concrete test of our trained model in real and synthesis image. We could the generated images are blurry; it is just because we diminished the quality of the images as this experiment run on a low-computational graphic card (RTX 2060).

Figure 4- Results of GANs on unseen example
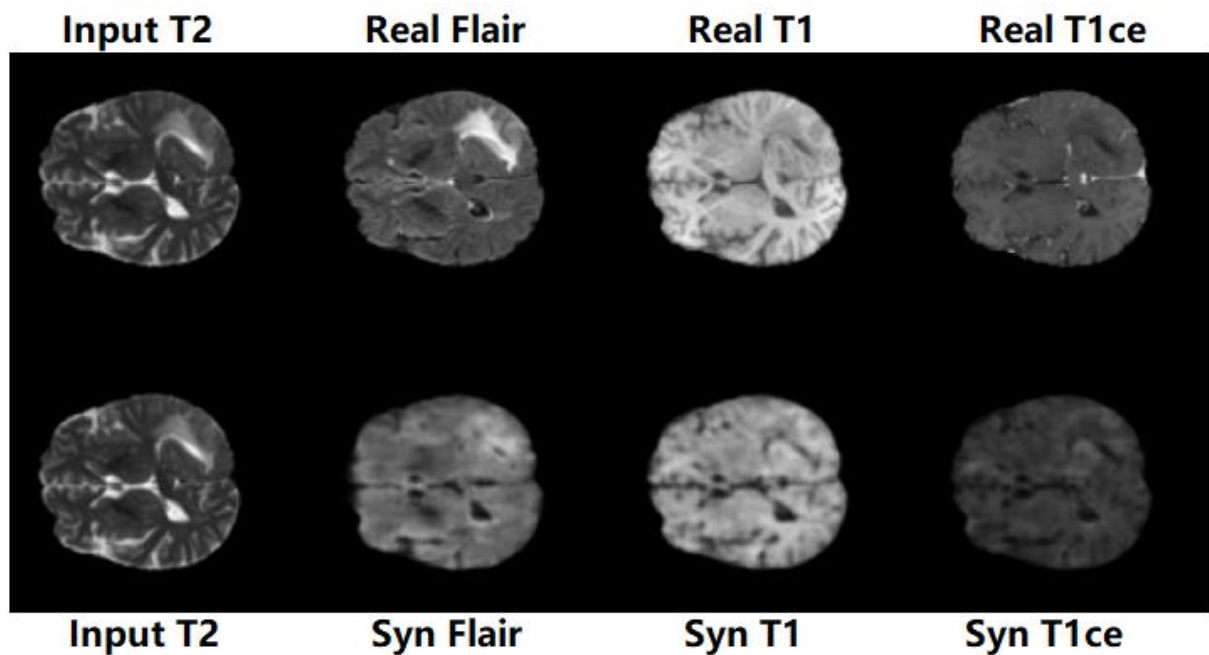
In the next step, we are going to enhance the results by running a full test of the GANs on the computational resource of the lab, using Tesla.

5. Conclusion

By the art of DL, we were able to synthesis 3 different modalities of MRI and that would save us much time and money in the mean of doing only one imaging of the patient. As we could use the T2 attribute, for example, to predict the other modalities. In this project, we have obtained a modest result on taking a sample of the BRATS2018.

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Choi, Y., Choi, M., Kim, M.S., Ha, J., Kim, S., & Choo, J. (2018). StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8789-8797.

[3] Dai, X., Lei, Y., Fu, Y., Curran, W.J., Liu, T., Mao, H., & Yang, X. (2020). Multimodal MRI Synthesis Using Unified Generative Adversarial Networks. Medical physics. https://doi.org/10.1002/mp.14539

[4] Menze et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE transactions on medical imaging, 34(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694

[5] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., & Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data, 4.

[6] Bakas et al. (2018). Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. ArXiv, abs/1811.02629.

[7] Xlim Présentation. (n.d.). Retrieved May 10, 2021, from https://www.xlim.fr/laboratoire/presentation

[8] Standford University. (2017). RSNA 2017: Rads who use AI will replace rads who don't. Retrieved May 3, 2021, from https://aimi.stanford.edu/news/rsna-2017-rads-who-use-ai-will-replace-rads-who-don-t

[9] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition.

[10] Meskó, B., & Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. npj Digit. Med. 3, 126. https://doi.org/10.1038/s41746-020-00333-z

[11] Krizhevsky, A., Sutskever, I., & G. E. Hinton. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages1097–1105.

[12] K. He, X. Zhang, S. Ren, & J. Sun. (2015). Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385.

[13] Isola, P., Zhu, J., Zhou, T., & Efros, A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5967-5976.

[14] Liu, M.-Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. arXiv preprint arXiv:1703.00848.

[15] Kim, T., Cha, M., Kim, H., Lee, J.K., & Kim, J. (2017). Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. ICML.

[16] Zhu, J., Park, T., Isola, P., & Efros, A.A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV), 2242-2251.

Annex

CelebA dataset contains 202,599 face images of celebrities, each annotated with 40 binary attributes. They crop the images, then resize them as 128×128. Randomly, they selected 2,000 images as test set, and they used all remaining images for training data. They constructed seven domains using the following attributes: hair color (black, blond, brown), gender (male/female), and age (young/old).

RaFD dataset consists of 4,824 images collected from 67 participants, who made eight facial expressions in three different gaze directions and angles. The images are cropped to 256×256, centralized faces, and finally resized to 128×128.