Reproducing the method of semi-supervised medical image segmentation via learning consistency under transformations on image classification of CIFAR-10 dataset

Wouter Polet
Tim Polderdijk

April 2021

1 Abstract

In this blog post we describe the reproduction of a simplified version of the semi-supervised medical image segmentation by Bortsova et al. (2019). We used the same approach to classify images of the CIFAR-10 dataset instead of the segmentation of the JSRT dataset. Additionally we tried using two different transformations on the dataset. Like the original paper, we found an increase of accuracy of the model using semi-supervised training instead of only supervised training.

2 Introduction

In the field of machine learning, and more specifically deep learning, a very important requirement to be able to train models with a relatively good performance is data. This data is traditionally used for supervised learning. The data includes labels, which is compared with the label given by the model. From this comparison a neural network computes some kind of error which is used to update the model. With enough input data a neural network should have 'learned' how to label inputs based on their data features. In general, the more training data you have, the better the performance of the resulting model will be. In this reproduction paper, we will also use unsupervised learning. A benefit of unsupervised learning is that it is easier and cheaper to obtain data without labels. This blog post is about a reproduction of Bortsova et al. (2019). That means we tried to reproduce the results they obtained, and to add value to the reproduction we used a new dataset, and made some small changes to the proposed method as well. The proposed method uses semi-supervised learning on a convolutional neural network to learn consistency under transformations when

performing image segmentation. This is done using a 'consitency loss' term in the loss function to minimize. By having a set of possible transformation tuples, batches of input data with a corresponding amount of transformations are made, and the loss function is approximated for each batch. The practical goal is that a model can be trained to correctly segment input images, even if they are transformed. It should not matter whether the transformation is linear or non-linear, as long as its inverse can be calculated it can be used. The original method was developed for x-ray images of the human chest, for medical purposes. In this reproduction we found that the method of learning consistency under transformations also works on a more general, public dataset (CIFAR-10 by Krizhevsky et al. (2009)). This dataset is a public dataset, widely used for image classification tasks. In the dataset there are 10 classes as can be seen in the figure below. The fact that it requires a relatively small amount of

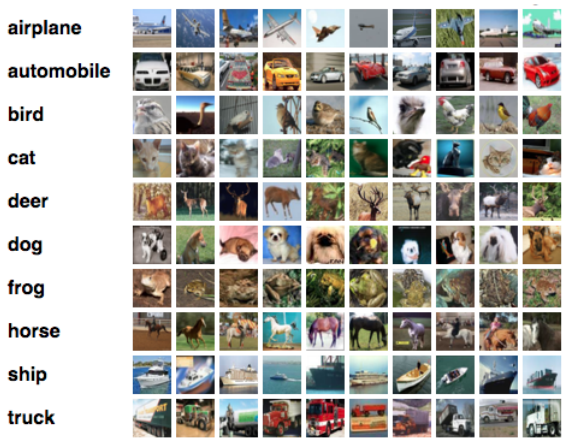

Figure 1: CIFAR-10 classes overview with some example images

labeled data, and is able to make use of semi-supervised learning to learn consitency under transformations gives the method practical value. The results of our reproduction show that the original method also works for training a CNN classifier on the CIFAR-10 dataset, not just for medical image segmentation on the JSRT dataset.

3 Related work

There are many works in the field of deep learning and image segmentation, and here we will focus on some prominent works about semi-supervised learning. Wang et al. (2019) proposes a framework called EnAET, which can be used to improve the performance of any semi supervised learning algorithm, and can also greatly improve the performance of supervised learning. The framework uses self-supervised learning to fully utilize the latent information in images,

and create a 'self-supervised term' which can be used in semi supervised and supervised learning algorithms. Closely related to the paper we reproduce, it aims to learn transformations, such that an image will still be recognized/classified correctly even if it is transformed.
In Li et al. (2018), a work that relates more to the medical application of semi supervised learning, semi supervised learning is used for image segmentation, to be able to detect skin lesion which is helpful when diagnosing patients. In this case semi supervised learning is used in stead of having an (expensive, inefficient) expert dermatologist make pixel-wise annotations. For their semi supervised method they use a transformation consistent scheme in their model, which is naturally helpful in practice as human skin is not a flat space and photographical consistency is often hard to find. Much like Bortsova et al. (2019) they also use and optimize a consistency loss function.
Another work which appeared quite prominently in our research is Sajjadi et al. (2016). It proposes a regularization method to improve the consistency over multiple passes of a deep neural network used in semi supervised learning. They also take into account that input images may be transformed (linearly or non-linearly) and should still be classified the same if they are passed through the network multiple times. The proposed method is an unsupervised loss function. They managed to obtain significant improvements in accuracy on multiple benchmark datasets when a small amount of labeled data was available.
There are other works as well, but we chose to feature these as they illustrate the idea of consistency under transformations and all use semi supervised learning. Sajjadi et al. (2016) and Wang et al. (2019) are more intended for general image classification, while Li et al. (2018) presents a specific case for image segmentation with a medical application.

4 Background

To understand how both the contribution of Bortsova et al. (2019) and our reproduction work, it is important to understand the concepts of deep learning and semi supervised learning.

4.1 Deep Learning

Deep learning allows machine learning models to learn feature representations of input data with multiple layers of abstraction LeCun et al. (2015). The key is that these layers are learned from data using general procedures, not designed by humans. As an example, for image classification, the first layer may detect the presence and orientation of edges. The second layer may detect certain arrangements of those edges. The third layer may detect parts of objects by those arrangements of edges. The fourth layer may detect objects from those parts, etc. For our reproduction we will use a Convolutional Neural Network (CNN), which is a type of neural network primarily used for image classification.

4.2 Semi Supervised Learning

There are 2 types of data that can be learned from, labeled and unlabeled data. Labeled data means that the data has a true label attached to it. In a classifier, this could mean we have an input image, and we know that it should be classified as a cat. But labeled data can be expensive to obtain or it might not be possible to obtain a sufficient amount of labeled data to train a machine learning network. At the same time, unlabeled data (data without a label attached to it) is usually cheaper and easier to obtain. Semi-supervised learning is a learning paradigm that allows models to make use of unlabeled data (given some labeled data, naturally) Zhu & Goldberg (2009). Semi-supervised learning means combining supervised learning (learning using the labels on labeled data) with unsupervised learning (learning using unlabeled data). In our reproduction, we used the same method of semi-supervised learning as Bortsova et al. (2019), but on a different dataset (CIFAR10).

4.3 Method Reproduced

The method that we reproduced defines a set of input images with their corresponding labels, and a set of images that are unlabeled. A distribution of tuples of mappings T is also defined. Each tuple contains two mappings (transformations), t^{in} and t^{out} such that the label corresponding to $t^{in}(x)$ is $t^{out}(y)$ where x is an input image, and y the corresponding label.

In order to have a neural network learn consistency under transformations, the parameters of the network should be optimized such that the following is minimal:
`regular supervised loss` + `unsupervised consistency loss`
The consistency loss term is what lets the network learn consistency under transformations. The proposed method attempts to optimize the parameters by applying a training scheme on small batches of training data. Each batch contains a set of labeled data, a set of unlabeled data, and two transformations from T. For each batch an objective function containing an approximation of both the supervised loss and consistency loss is minimized.
A key point to keep in mind is that the method works with any transformation of which the inverse of t^{out} is computable.

5 Reproduction setup

Bortsova et al. (2019) used the JSRT dataset, which is a dataset filled with s-ray images of human chests. In this reproduction, we applied the technique on a different dataset; the CIFAR-10 dataset. This dataset contains 32x32 pixels images that can be divided into 10 different classes. The network should then decide which class an image belongs to. The original dataset consists of 60,000 images, but we only use 4,000 labeled images. In the unsupervised step, we add 45,000 unlabeled (transformed) images. Because this problem is a classification

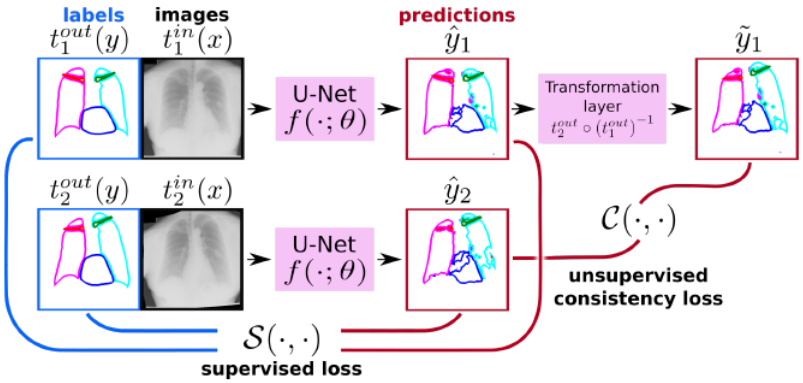

Figure 2: Overview of the original method architecture

problem, rather than a segmentation problem, we only need to transform the input.

An example of some random rotation transformations can be seen in the figure below.

Figure 3: 3 Randomly rotated input images concatenated

Besides the new dataset, we also altered the method slightly. Instead of the elastic deform used in Bortsova et al. (2019), we opted for a rotation over a random angle. We varied the range of this angle, to find the best range. Besides this random rotation, we also used an elastic deformation, as the paper does. This deformation shifts the image by a random amount of pixels.

The training of the model consists of two steps: supervised and unsupervised learning. In the first step, we train the model on labeled images only. We train a small convolutional neural network on the 4000 labeled images for 128 epochs, with a batch size of 128, and an Adam optimizer with learning rate of 0.001. Then we pick the solution with the lowest loss and save it.

The second step continues with the solution picked in the first step and continues training it. The loss now becomes the average of the supervised and unsupervised losses. Furthermore we lower the learning rate to 0.0005. With the higher learning rate, the steps taken are too big. With the unlabeled (transformed) images, this would only end the training step with a lower accuracy rate. In this step we train at most 128 epochs, but stop whenever the loss did not improve for 10 epochs.

The goal of this experiment is to find whether the unsupervised learning step improves the accuracy of the network, like was shown in Bortsova et al. (2019). The code used for this reproduction is a modified template that can be found at: `https://github.com/Gerda92/pyoneer`. The code with its modifications can be found here: `https://github.com/wouterpolet/pyoneer`.

6 Results

We start with training a baseline model in a supervised manner. Figure 4 shows the validation loss of this training.

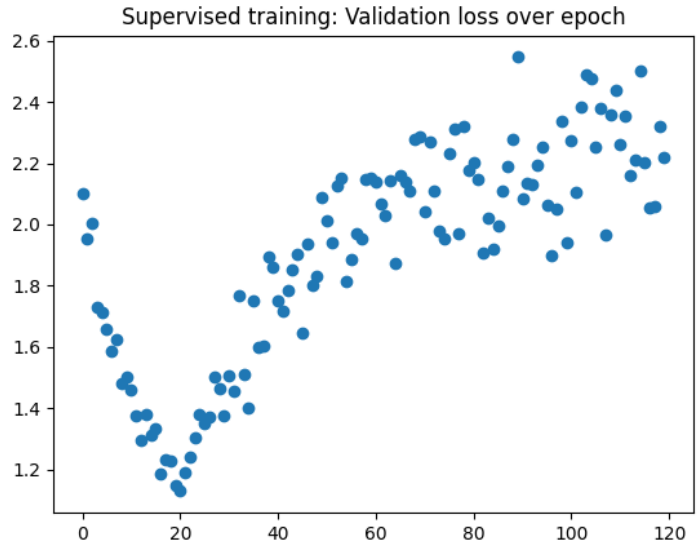

Figure 4: Validation loss over epochs of the training of the base model.

Taking the model with the lowest loss (at 20 epochs) we get an accuracy on the test data of 62%.

Using the random rotation, we were able to increase the accuracy of the model to 73%. Searching for a sensible range for the angles of the rotation, we trained the model with different ranges. The validation losses are shown in figure 5. Here a range of 1 means that the rotation can range from $-\pi$ to π. A range of 0.5 would mean from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$. The model with the best result happened to have a range of 0.05.

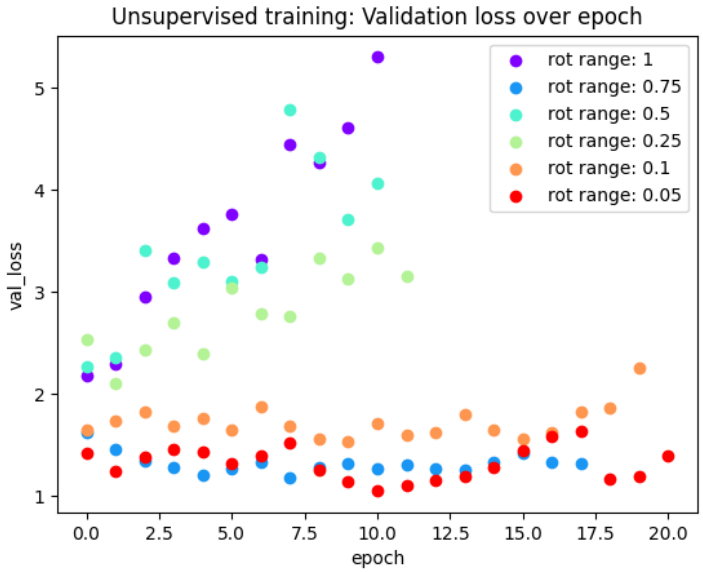

Figure 5: Validation loss over epoch for different ranges of the angle of rotation

We cannot find a clear relation between the range and the validation losses. It seems that the range itself does not change much, but some trainings happen to work better than others. This does not make using the rotation reliable, but it is possible to achieve a better trained model.

Using the random shift transformation, we found a similar increase in accuracy; to 77% in the best case. This was achieved with a max shift of 5. Again we trained the model with various values for the maximum shift. The results are shown in figure 6.

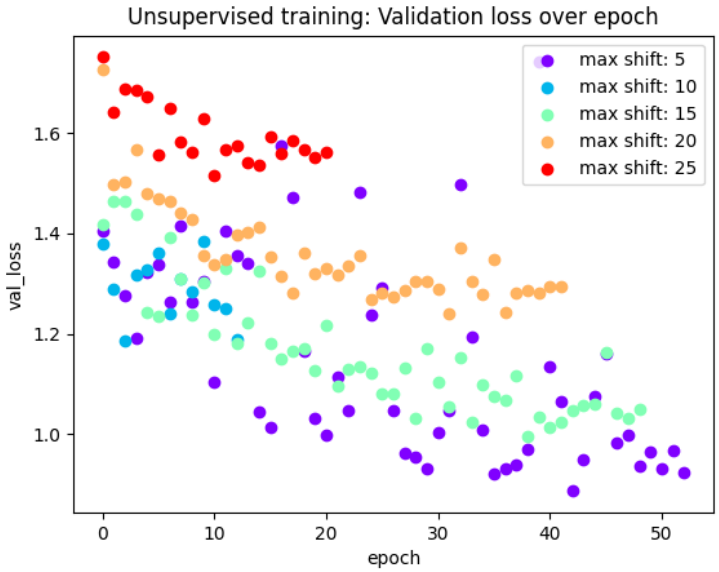

Figure 6: Validation loss over epoch for different maximum shift

We can see that the loss decreases more consistently than with the rotation. The higher the maximum shift, the worse the model seems to perform. This is likely due to the large black space that is created by a large shift. With only 32 by 32 pixels in an image, shifting it 15 pixels can already remove half of the image. We see that, in our tests, the maximum shift of 5 performs best. In general the random shift performs better than the random rotation transformation.

7 Conclusion

This is a blog post about the reproduction of a simplified version of the method of learning consistency under transformations proposed by Bortsova et al. (2019). This method includes a consistency loss term in the loss function when training a neural network, and separates data in batches after which the loss function is approximated for each batch. The method has practical value because it uses semi-supervised learning, and consistency under transformations is important for a robust image classification or segmentation model in practice.
We managed to reproduce the results when using the method to train a CNN on the CIFAR-10 dataset to show that the learning of consistency under transformations works. Different types of transformations were tried, non-linear elastic deformations and random angle rotations. The results show that the method does indeed seem to work as the proposed method increased the accuracy of

our model from 62% to 73%. There does not seem to be a clear relation between the rotation angle range and the validation loss. When using random shift transformations (elastic deformations) the accuracy was increased to 77% in the best case. A relatively low maximum shift value of 5 seems to be optimal in our setup. According to the reproduced results, the original method can thus be used in fields other than medical image segmentation as well to improve the performance of a neural network by learning consistency under transformations.

References

Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., & de Bruijne, M. (2019). Semi-supervised medical image segmentation via learning consistency under transformations. In *International conference on medical image computing and computer-assisted intervention* (pp. 810–818).

Krizhevsky, A., Hinton, G., et al. (2009). Learning multiple layers of features from tiny images.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Li, X., Yu, L., Chen, H., Fu, C.-W., & Heng, P.-A. (2018). Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. *arXiv preprint arXiv:1808.03887*.

Sajjadi, M., Javanmardi, M., & Tasdizen, T. (2016). Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *arXiv preprint arXiv:1606.04586*.

Wang, X., Kihara, D., Luo, J., & Qi, G.-J. (2019). Enaet: Self-trained ensemble autoencoding transformations for semi-supervised learning. *arXiv preprint arXiv:1911.09265*.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, *3*(1), 1–130.