# Research statement and proposal

**Ying Peng**

*1853287@tongji.edu.cn*

## 1 Research statement

AlexNet, proposed in 2012, first applied convolutional neural network in image classification task. Since then, computer vision has boosted in recent 10 years. R-CNN, appeared in 2012, brought deep learning based object detection method into public view. YOLO, put forward in 2014, the first unified, real-time one-stage object detector whose fast response makes the development of autonomous driving become possible, has developed several versions in these years and been applied in practice of production and life. Machine learning methods improve dramatically and are widely used in industry such as self-driving cars, medical diagnosis, automation and robotics. Researchers keep on pushing the extension of machine's consciousness and ability. We hope that machines will do most of the work for us in the near future.

## 2 Past and future research

My first research project focuses on meta-learning and attention mechanism in complex domain for signal recognition. For these tasks, a large volume of data is required to obtain satisfactory performance. However, the deep learning models trained with traditional supervised learning methods often perform poorly or even fail when only a small amount of data is available or when they need to adapt to unseen tasks or time-varying ones. The collection and annotation of abundant data are notoriously expensive, especially for some rare but important signals. Another critical challenge is the presence of noise because the signal data varies for different signal-to-noise ratios (SNRs) and the deep neural networks (DNNs) have to adapt to real-time variations in SNRs.

Because, in the real world, signal annotation is expensive and models need to adapt to changing SNRs, whereas meta-learning has an explicit goal of fast adaptation which seeks to resolve above challenges by learning how to learn like humans do. The meta-learner is trained on the distribution of homogeneous tasks, with the goal of learning internal features that are broadly applicable to all tasks, rather than a single individual task. Additionally, Since the signals contain both magnitude and phase, complex numbers are used for the representation of signals. Inspired by the special structure of the signal, i.e., real and imaginary parts consisted in practical time-series signals, complex-valued neural networks and attention could be applied to take advantage of the prior knowledge of the signals. We refer to the complex-valued neural networks that define complex derivatives and the complex chain rules as complex derivatives complex-valued neural networks(CDCVNN). The complex chain rule is a key to implementing CDCVNNs. According to the complex chain rule, we are able to derive the outer-loop update process of CAMEL, which is different from that of MAML. In this way we are able to design complex-valued meta-learning.

We adopted multi-head attention to get better performance and parallel computation. To implement Complex-valued Attention, we define the Complex gradient vector and Complex-valued SoftMax function, at last the complex-valued multi-head attention, which allows models to jointly focus on information from different representations. Complex-valued normalization and Complex-valued activation functions are also taken into account to build the complex-valued neural network. We derive and prove the convergence analysis of complex-valued MAML, which is shown that CAMEL is able to find an $\varepsilon$ first-order stationary point for any positive $\varepsilon$ after at most $O(1/\varepsilon^2)$ iterations with second-order information. With second-order information, CAMEL is able to find first-order stationary points of general nonconvex problems.

The work was done in collaboration with professor Qingjiang Shi from Tongji University. I designed all the code, finished the whole experiment by myself and wrote the main part of the paper. I'm now applying to other types of datasets like videos, music and time series. The paper is available at https://arxiv.org/abs/2106.04392. I have applied a patent for this work and it is under review.

My second research experience is taking part in the Kaggle Research Prediction

Competition: SETI Breakthrough Listen - E.T. Signal Search. It took raw data from the telescope and performed a Fourier Transform to generate a spectrogram as an input sample. These spectrograms consist of measurements of signal intensity as a function of frequency and time, accompanied by headers containing metadata such as the direction the telescope was pointed in, the frequency scale, and so on. As searching for candidate signatures of extraterrestrial technology, the main obstacle to doing so is that our own human technology also gives off radio signals, referred as “radio frequency interference”, or RFI. One method we use to isolate candidate technosignatures from RFI is to look for signals that appear to be coming from particular positions on the sky. Typically, we do this by alternating observations of our primary target star (A) with observations of three nearby stars (B, C, D). One set of six observations (ABACAD) is referred to as a “cadence”. For positive samples, there are “needle” signals in “A” observations. Not all of the “needle” signals look like diagonal lines, and they may not be present for the entirety of all three “A” observations, but what they do have in common is that they are only present in some or all of the “A” observations. Participants are required to design a model to figure out potential E.T. signals. (See details at https://www.kaggle.com/c/seti-breakthrough-listen/)

My main work in this competition includes performing data processing, model selection, training, and optimization. My group got excellent accuracy through model ensemble and pseudo label learning and gained silver medal in this competition. (Ranking: 34/768). Main methods and tricks we considered are as follows:

**Data Processing** Firstly, I considered the sample-signal in frequency, as a greyscale image. Secondly, I concatenated 6 channels vertically as one picture. In addition, I have tried to input as a 6-channel image sample but got worse performance than concatenation. What’s more, the loss of information because of concatenation need considered: the origin 6 subsamples are regarded as one. To let our model know that there were 6 parts, we segmented 6 parts by drawing 5 lines manually.

**Data Augmentation** Samples are very uneven in the dataset that the number of negative samples is much larger than the positive, so we oversampled the negative samples. What’s more, we also oversampled the hard sample after 2 folds. Another trick we used is drawing some diagonal lines to simulate RFI and made it harder to recognize technosignatures so that the model would be better-trained and have satisfactory performance. But after the competition we learned that in the datasets there are samples with curves and if we had drawn curves, we would get results even better, so I learned to observe the samples more detailed before designing solutions after that.

**Model Ensemble** We mainly used efficientnet_b0 and resnet_34d 2 models. We modified the parameters and trained models for 5 folds to get 5 models each time and selected models with good performance to ensemble. Reading other teams' solutions, we found out using bigger models like efficientnet_b5 may have better performance. We need to consider the tradeoff between time, computation and accuracy and choose better models next time.

**Unsupervised Learning** We tested our trained model and set a confidence threshold. samples with results larger than the threshold would become training data with pseudo labels for next steps. In this way the whole process is: 1. Train a bunch of models and select models with good performance. (Use the validation set score to judge the performance of the model.) 2. Test these well-performing models on testing set and compute the mean of the results as the output label. 3. Train a new model using training set and samples with high confidence pseudo labels.

Other works include Detection of Diabetic Retinopathy based on Deep Learning (Cooperation with Tongji Hospital). I established the model based on the Unet and transfer learning to extract image features, optimized the loss function of the classifier by square-weighted Kappa to learn the order information between categories, filtered out images with good quality for classification through the quality classification model of fundus image to further improve the accuracy of the model and built a front page to form a complete system that can be used. I took a Deep Learning Framework Development Intern in Sensetime Research Institute. I Participated in developing the deep learning framework-Parrots, including neural network operators developing, model precision commissioning, adaptive quantitative training, mixed precision training and precision analysis visualization tool development.

## 3 Motivation for future research

Stand-alone self-attention layer proposed in 2019 greatly reduces the amount of parameters and calculation with similar model accuracy compared with traditional convolution models in image classification and target detection tasks. Adversarial data restricts the further application of machine learning in areas where safety is vital such as self-driving cars, medical diagnosis, financial analysis and so on, so it is of great importance to improve the adversarial robustness of neural networks. Other fields like reinforcement learning and manifold learning are also hot topics. I keep learning related knowledge and reading Influential paper. Furthermore, I take part in research and projects using machine learning to solve problems in reality. Through these experiences I get familiar with state-of-art networks and techniques and apply them in real scenes. I hope to go on further study in machine learning related fields, for example, computer vision, robotics, geometric modeling, etc. I will be devoted to exploring in this innovative field and working hard on applying techniques in industries, utilizing knowledge I have to truly benefit daily life.