Contents

[Study plan for exploring Position Trainable Transformers for Computer Vision 1](#_Toc179466901)

[Month 1-x0: Foundations of Transformers and Computer Vision 1](#_Toc179466902)

[Month x0-x1: Vision Transformers (ViTs) 2](#_Toc179466903)

[Month x1-x2: Exploring Position Trainable Transformers 2](#_Toc179466904)

[Month x2-x3: Applications of Transformers in Computer Vision 3](#_Toc179466905)

[Month x2-x3: Self-Supervised Learning and Transformers (In case you are interested) 3](#_Toc179466906)

[Month x3-x4: Experimentation and Final Project 3](#_Toc179466907)

# Study plan for exploring Position Trainable Transformers for Computer Vision

## Month 1-x0: Foundations of Transformers and Computer Vision

**- Week 0-y0**: Transformers Overview

- Readings:

* "Attention Is All You Need" by Vaswani et al. (2017). (READ, READ, READ)
* A primer on self-attention mechanisms.

- Videos/Courses:

* Such as — Stanford's CS231n lectures on transformers (lecture 9)—

**- Milestone 1**:

- Experiment with simple transformer architectures using PyTorch or TensorFlow (Preferably PyTorch).

**- Week y0-y1**: Computer Vision with Deep Learning

- Readings:

* Goodfellow et al.'s “Deep Learning” (focus on CNNs and vision-specific chapters).

**- Milestone 2**:

* Implement basic CNN architectures like ResNet and VGG for image classification tasks on simple available datasets like CIFAR-10.

**- Project**:

* Use a pre-trained vision transformer (ViT) on a simple dataset (ImageNet or CIFAR-10) and evaluate its performance.

## Month x0-x1: Vision Transformers (ViTs)

**- Week y1-y2**: ViT Basics

- Readings:

* "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" by Dosovitskiy et al. (2020).

- Videos/Courses:

* Hugging Face Course on Vision Transformers.

**- Milestone 3**:

* Train a ViT from scratch on a small custom dataset (e.g., MNIST, CIFAR-10 or Neadvance may provide a dataset until then).

**- Week y3-y4**: Advanced ViT

- Readings:

* "Scaling Vision Transformers" by Zhai et al. (2021).
* "Exploring Simple Siamese Representation Learning" (related to self-supervision).

**- Milestone 4**:

* Fine-tune a pre-trained ViT model on more complex datasets, such as ImageNet.

## Month x1-x2: Exploring Position Trainable Transformers

**- Week y4-y5**: Positional Embeddings

- Readings:

* Study how positional encodings work in transformers.
* Papers like "Understanding Positional Encodings in Transformers" (from different research).

**- Milestone 5**:

* Experiment with modifying the positional embeddings of a ViT model (train with different configurations).

**- Week y5-y6**: Position Trainable Transformers (PTT)

- Readings:

* Papers related to Position Trainable Transformers or Dynamic Positional Encodings.

**- Milestone 6**:

* Implement position trainable modules and apply them to standard transformer architectures.

**- Project**:

* Compare the results of static vs. position trainable transformers on a custom dataset.

## Month x2-x3: Applications of Transformers in Computer Vision

**- Week y6-y7**: Transformer Applications

- Readings:

* "End-to-End Object Detection with Transformers" (DETR).
* "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows".

**- Milestone**7:

* Implement DETR and Swin Transformer models for object detection tasks.

**- Week y7-y8**: Multimodal Transformers

- Readings:

* "Image-Text Pre-training for Vision and Language" (CLIP, ALIGN).

**- Milestone 8**:

* Use CLIP models for image-text understanding tasks and fine-tune it for specific use cases.

## Month x2-x3: Self-Supervised Learning and Transformers (In case you are interested)

**- Week y8-y9**: Self-Supervised Learning Basics

- Readings:

* Papers on self-supervised learning such as "SimCLR" and "BYOL".

**- Milestone 9**:

* Implement a self-supervised learning framework using vision transformers.

## Month x3-x4: Experimentation and Final Project

**- Week y9-y10**: Research and Experimentation

* Start developing a final research project on Position Trainable Transformers in a specific domain (e.g., object detection, segmentation, hyperspectral imaging).
* Formulate a hypothesis, plan experiments, and define evaluation metrics.

**- Week y11-y12**: Writing and Reporting

- Prepare a detailed report on your findings, comparing performance of position trainable transformers with conventional transformers across various tasks.

* Write a research paper based on your project with experimental results (—optional—).

By the end of this study plan, you should have a deep understanding of Position Trainable Transformers in the context of computer vision, be able to implement various transformer architectures, and contribute to research in this field.

*ALSO, KEEP IN YOUR MIND REGARDING NEADVANCE’ REGULARIZATIONS WE EXPECT A COMPLETE COMPREHENSIVE REPORT (E.G. POWERPOINT OR A WORD FILE) AFTER EACH MILESTONE.*