Mini_project 2+3 Description

In mini-projects 2 and 3, you are expected to pick a deep learning problem of your own choosing — ***our only requirement is that the project has an element of novelty***, that is, you have to advance the state-of-art in some way. For example, your project could build on a published paper (or a combination of papers) from a machine learning conference (*examples*: include https://nips.cc/; https://icml.cc/; https://www.aaai.org/ amongst others). Your project does not have to be based on published paper(s) however; you could also, for example, look at new datasets for which there hasn't been much work (*example*: https://nips.cc/Conferences/2021/DatasetsBenchmarks/AcceptedPapers).

Ideally, when picking a project, you want to consider the following factors:

- **Are datasets and benchmarks readily available**? Collecting/curating new datasets is a very time-consuming task and will likely not be possible in the time-frame of a class.
- **Is there existing code you can build on**? Most papers these days link to a GitHub repo. Check to see if they provide clear instructions on running and reproducing the code and results. Also check to see if the project is “popular” for example based on the number of downloads, ratings etc. This is a good sign that the code works.
- **Is there a way you can identify to extend the paper or method**? The extension doesn’t have to be novel in the sense that it might lead to a new paper at an ML conference (although that would be cool!) — for example, you could evaluate the method on a new benchmark, on a new or different architecture, or try and improve the performance of the method. The paper’s future work section might help.
- **What about training time?** Some architectures and/or datasets are prohibitively time consuming to train; you cannot reasonably train such models in the time that we have. The first thing you should do is make sure that the project can be completed in a reasonable time. For example, you can try your idea on a smaller dataset, transfer from a pre-trained model, or fine-tune models instead of training from scratch. But bottom line: beware of training time. And start early!

These are the key milestones for your mini-project:

1. **Project Proposal Submission (Due Friday Midnight April 1st):** An up-to two page document with: (1) Project title; (2) Team members; (3) Key idea and datasets: describe what you plan to do along with references of papers you will build on, the datasets you plan to use, and ; (4) Deliverables: identify deliverables, that is, what you will deliver if your project is successful (example: a new)
2. **Feedback on proposal:** The teaching team will provide feedback on your project proposal by Friday April 8th, with a focus on what needs to be changed. For example, we may say your task is too ambitious and ask you to try fewer things, or conversely, that it needs to be more ambitious.

3. **Mid-Project Report (replaces mini-project 2, due Fri April 22nd Midnight):** A 6-page report in Neurips format describing your progress so far. By this stage, we require you have atleast some working code, and an initial/preliminary set of results.
4. **Final Report (replaces mini-project 3, due Tue May 17th Midnight):** An 8-page report in Neurips format describing your project and its outcomes.

You will find below some papers selected by your teaching staff. Note that these are just some of our favourite recent papers, and ***by no means*** reflect that we want you to do projects based on these papers.

—---

Intriguing Properties of Vision Transformers

Paper Link: https://openreview.net/forum?id=o2mbl-Hmfgd
Code Link: https://github.com/Muzammal-Naseer/Intriguing-Properties-of-Vision-Transformers

Summary: Transformer architectures have achieved SOTA for a range of tasks. This paper seeks to dig deeper into the properties of transformer architectures, for example, do they focus more on shape or on texture cues for image tasks, or how robust transformers are to certain types of “perturbations” for example, if certain parts of an image are blocked out. The paper also proposes a new method for training “shape-biased” transformers.

Revisiting Contrastive Methods for Unsupervised Learning of Visual Representations

Paper Link: https://openreview.net/pdf?id=j2gshvolULz
Code Link: https://github.com/wvangansbeke/Revisiting-Contrastive-SSL

Short Summary: *Contrastive learning is one of the most exciting advances in semi-supervised learning*. The paper tries to study the impact of different datasets and augmentation strategies for contrastive learning. New augmentation strategies that can improve contrastive learning are also proposed.

Lossy compression for lossless prediction
Paper Link: https://nips.cc/virtual/2021/spotlight/26845
Code Link: https://github.com/YannDubs/lossyless

A Closed-form Solution to Photorealistic Image Stylization
Paper Link:
https://www.ecva.net/papers/eccv_2018/papers_ECCV/papers/Yijun_Li_A_Closed-form_Solution_ECCV_2018_paper.pdf
Code Link: https://github.com/NVIDIA/FastPhotoStyle

Short summary: Photorealistic image stylization concerns transferring style of a reference photo to a content photo with the constraint that the stylized photo should remain photorealistic. While several photorealistic image stylization methods exist, they tend to generate spatially inconsistent stylizations with noticeable artifacts. This paper addresses these issues, while conducting extensive experimental validations.

GANs For Face Swapping

Paper Link:

1. https://paperswithcode.com/paper/deepfacelab-a-simple-flexible-and-extensible
2. https://arxiv.org/pdf/1908.05932.pdf

Code Link:

1. https://github.com/IanSullivan/DeepFakeTorch
2. https://github.com/YuvalNirkin/fsgan

Short summary: These papers seek to use GANs to build a face swapping program with better performance than existing code. Paper 1 details the pipeline of DeepFaceLab, a framework that is easy to use and can produce high quality face swapping. Paper 2 details the structure of a FSGAN - an RNN based approach that seeks to improve the face swapping algorithm by addressing issues like pose and expression variations, seamless interpolation for video feeds and the preservation of target skin color and lighting conditions.

Conditional Generative Adversarial Nets

Paper Link: https://arxiv.org/pdf/1411.1784.pdf

Code Link: https://github.com/arturml/mnist-cgan

Abstract : Generative Adversarial Nets were recently introduced as a novel way to train generative models. In this work we introduce the conditional version of generative adversarial nets, which can be constructed by simply feeding the data, y, we wish to condition on both the generator and discriminator. We show that this model can generate MNIST digits conditioned on class labels. We also illustrate how this model could be used to learn a multi-modal model, and *provide preliminary examples of an application to image tagging in which we demonstrate how this approach can generate descriptive tags which are not part of training labels.*

Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers

Paper: https://arxiv.org/pdf/2012.15840.pdf
Github: https://github.com/fudan-zvg/SETR

Most recent semantic segmentation methods adopt a fully-convolutional network (FCN) with an encoder/decoder architecture. The encoder progressively reduces the spatial resolution and learns more abstract/semantic visual concepts with larger receptive fields. Since context modeling is critical for segmentation, the latest efforts have been focused on increasing the receptive field, through either dilated/atrous convolutions or inserting attention modules. However, the encoder-decoder based FCN architecture remains unchanged. In this paper, we aim to provide an alternative perspective by treating semantic segmentation as a sequence-to-sequence prediction task. Specifically, we deploy a pure transformer (i.e., without convolution and resolution reduction) to encode an image as a sequence of patches. With the global context modeled in every layer of the transformer, this encoder can be combined with a simple decoder to provide a powerful segmentation model, termed SEgmentation TRansformer (SETR). Extensive experiments show that SETR achieves new state of the art on ADE20K (50.28% mIoU), Pascal Context (55.83% mIoU) and competitive results on Cityscapes. Particularly, we achieve the first position in the highly competitive ADE20K test server leaderboard on the day of submission.

—---

Linformer: Self-Attention with Linear Complexity

Paper: https://arxiv.org/pdf/2006.04768.pdf
Github: https://github.com/lucidrains/linformer
How its used:
https://ai.facebook.com/blog/how-facebook-uses-super-efficient-ai-models-to-detect-hate-speech/

Large transformer models have shown extraordinary success in achieving state-ofthe-art results in many natural language processing applications. However, training and deploying these models can be prohibitively costly for long sequences, as the standard self-attention mechanism of the Transformer uses O(n 2) time and space with respect to sequence length. In this paper, we demonstrate that the self-attention mechanism can be approximated by a low-rank matrix. We further exploit this finding to propose a new self-attention mechanism, which reduces the overall self-attention complexity from O(n 2) to O(n) in both time and space. The resulting linear transformer, the Linformer, performs on par with standard Transformer models, while being much more memory- and time-efficient

—---

MLP-Mixer: An all-MLP Architecture for Vision

Convolutional Neural Networks (CNNs) are the go-to model for computer vision. Recently, attention-based networks, such as the Vision Transformer, have also become popular. In this paper we show that while convolutions and attention are both sufficient for good performance, neither of them are necessary. We present MLP-Mixer, an architecture based exclusively on multi-layer perceptrons (MLPs). MLP-Mixer contains two types of layers: one with MLPs applied independently to image patches (i.e. “mixing” the per-location features), and one with MLPs applied across patches (i.e. “mixing” spatial information). When trained on large datasets, or with modern regularization schemes, MLP-Mixer attains competitive scores on image classification benchmarks, with pre-training and inference cost comparable to state-of-the-art models. We hope that these results spark further research beyond the realms of well established CNNs and Transformers.