아주대학교
세상의 A+가 되겠습니다

2023-2
AJOU SOFTCON

An Initial Study of Deep Learning Architecture for Image Recognition

이 름 서동건

지도교수 유종빈

연구배경

- 딥러닝 기술을 활용한 컴퓨터비전 분야에 관심과 흥미가 생겨 깊고 전문적인 공부와 연구를 진행하고자 유종빈 교수님과 함께 하는 자기주도연구에 참여하게 되었다.
- 기초 학습을 통해 연구를 수행할 수 있는 준비를 하는 단계를 거치고 나면, 선행 연구를 조사하고 이해하는 과정과 함께 직접 구현하는 과정을 통해 딥러닝 아키텍처의 높은 이해를 도모하고자 한다. 마지막으로 실제 현업과 연구의 최전선에서 이루어지고 있는 양상을 이해하고 구현하는 것이 목표이다.

결과 및 분석

- 본격적인 연구에 들어가기 앞서, 연구에 기반이 되는 개발환경 세팅을 진행했다. 이후 최신 Vision transformer 와 Convolutional neural networks 기반의 아키텍처에 대한 이해를 먼저 진행하여 통찰력을 다지고 들어가고자 했다. 따라서 주마다 3~4 편의 Paper Study 를 진행하며 기초 이론에 대한 학습을 진행했다. 이와 같은 기초 학습을 통해 연구를 수행할 수 있는 준비를 하는 단계를 거쳤다.

Week	Paper	Conf	Year
1	ImageNet Classification with Deep Convolutional Neural Networks	NIPs	2017
	Very Deep Convolutional Networks for Large-Scale Image Recognition	Arxiv	2014
	Deep Residual Learning for Image Recognition	CVPR	2016
	CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features	ICCV	2019
2	Squeeze-and-Excitation Networks	CVPR	2018
	CBAM: Convolutional Block Attention Module	ECCV	2018
	EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks	ICML	2019
	Distilling the Knowledge in a Neural Network	NIPs	2014
3	mixup: Beyond Empirical Risk Minimization	ICLR	2018
	Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift	ICML	2015
	RandAugment: Practical automated data augmentation with a reduced search space	CVPR	2020
4	ResNet strikes back: An improved training procedure in timm	Arxiv	2021
	AutoAugment: Learning Augmentation Policies from Data	CVPR	2019
	Random Erasing Data Augmentation	AAAI	2020
5 & 6	Rich feature hierarchies for accurate object detection and semantic segmentation	CVPR	2014
	Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks	NIPs	2015
	You Only Look Once: Unified, Real-Time Object Detection	CVPR	2016
	SSD: Single Shot MultiBox Detector	ECCV	2016
	Feature Pyramid Networks for Object Detection	CVPR	2017
7	An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale	ICLR	2021
	Methods for interpreting and understanding deep neural networks	DSP	2018
	Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization	ICCV	2017
8	Intriguing Properties of Vision Transformers	NIPs	2021
	ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness	ICLR	2019
	How Do Vision Transformers Work?	ICLR	2022

- 뿐만 아니라 직접 구현하는 실습과정을 통해 딥러닝 아키텍처의 높은 이해를 가지고 실제 현업과 연구의 최전선에서 이루어지고 있는 양상을 이해하고 구현해 보았다. 이후 실제로 진행한 연구를 표와 DB 로 정리해보고 결과값에 대한 생각과 고찰의 시간을 가져 향후 연구에 대한 연구방향성을 확립했다.
- 이 결과값들을 단순히 표와 DB 에서 그치지 않고, 실험과 연구를 효율적이고 신뢰성 있게 설득하기 위해 성능 향상 근거나 논리적인 설명에 사용하는 다양한 Analysis 기법들을 적용하여 나타내고 표현해 보며 Analysis의 중요성을 체감했다.

- 이를 실제로 Fourier frequency Analysis기법을 사용하여 resnet과 dino를 비교분석해 CNN과 ViT의 차이점을 나타내 보았고, 아래와 같은 결론을 도출했다.

Architecture	frequency	Vs human	Texture vs shape
CNN(resnet)	high frequency를 많이 봄	사람과 반대(이미지에서 textur에 집중)	high frequency를 많이 보니 texture를 많이 보는 것 증명
ViT(dino)	low frequency를 많이 봄	사람과 유사(이미지에서 shape에 집중)	low frequency를 많이 보니 shape를 많이 보는 것 증명

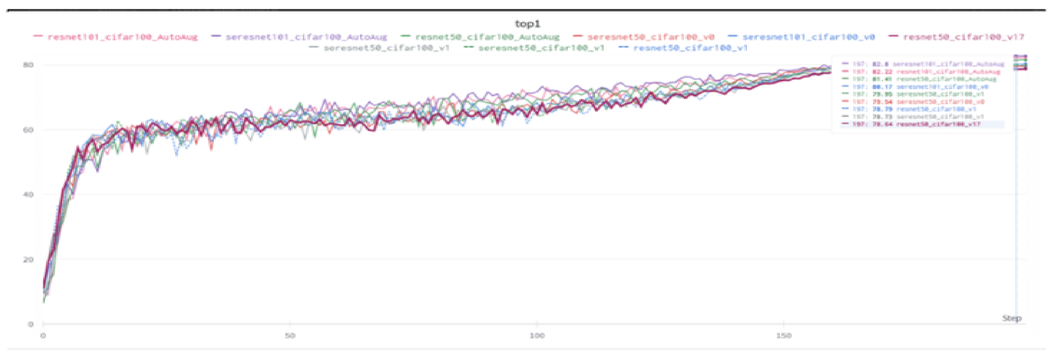
연구진행과정

1. On CIFAR100 dataset, achieves top-1 accuracy > 82% within 2 train hour

이번 연구를 통해 data augmentation 을 이용한 데이터 처리 해 방법이 1~2%의 성능향상이 나타나는 것을 확인해 보았다. 뿐만 아니라 논문에서 학습했던 SE 를 모델에 적용하고 성능이 1~2%정도 향상하는 것을 볼 수 있었다. Model Architecture 의 깊이를 더욱 깊게 만들어보는 시도도 해보았는데, 이 경우에는 성능향상이 일어날지는 몰라도 파라미터의 수와 Train Time 이 현저하게 증가하는 것을 확인했다. 따라서 깊이를 조절하는 방법은 주의를 하며 사용해야 하겠다 라는 결론을 내렸다.

The experiment results

Dataset	Model	Top-1 Acc	Top-5 Acc	Params	Train Time
cifar100	ResNet50	79.05	93.94	23M	1h 7m 26s
cifar100	SENet50	80.00	94.01	23M	1h 8m 0s
cifar100	ResNet50 + AutoAugment	81.62	94.94	23M	1h 6m 48s
cifar100	SENet101 + AutoAugment	82.85	95.73	47M	2h 25m 40s



2. On CIFAR100 dataset, achieves top-1 accuracy > 82% using timm library's various augmentation techniques. only use the given model(ResNet-50)

이번 연구를 통해 세세한 값들의 augmented들을 argparse 기법을 사용하여 세세하게 파인튜닝을 진행하며 최적의, 최고의 결과값을 찾아 나섰지만 완벽하게 목표에 달성한 실험결과는 아쉽게도 존재하지 않았다. 하지만 cifar100 dataset에서 Resnet50 Model로 image size를 224 x 224로 키운 결과가 81.91로 가장 목표에 근접한 결과값으로 나타났다.

따라서 image size를 적절한 크기로 Increasing 시키고 randomcrop과 autoaugment 방법을 사용하면 acacuracy 향상에 도움이 된다는 것을 확인했다.

The experiment results

Dataset	Model	Optimizer/ batch_size	augmentation techniques	GPU-개수	Epoch	Top-1 Acc	Train Time
cifar100	ResNet50	sgd / 256	mean,std 0.5 / cutmix 1 / lr 1 / warmup_lr 1e-4 / colorjiter 0 / randomcrop 사용 / autoaugment	DDP - 2개	200	61.39	20m 13s
cifar100	ResNet50	sgd / 256	mean,std 0.5 / cutmix 1 / lr 1 / warmup_lr 1e-4 / colorjiter 0 / randomcrop 사용 / autoaugment / momentum 0.9 / weight-decay 1e-4	DDP - 2개	200	67.78	23m 56s
cifar100	ResNet50	sgd / 256	mean,std 0.5 / cutmix 1 / lr 1 / warmup_lr 1e-4 / colorjiter 0.4 / randomcrop 사용 / autoaugment v0 / momentum 0.9 / weight-decay 1e-4 / decay-rate0.1 / smoothing0.1	DDP - 2개	200	67.64	23m 47s
cifar100	ResNet50 _tutorial	sgd / 256	mean,std 0.5 / cutmix 1 / lr 0.5 / warmup_lr 1e-4 / colorjiter 0 / randomcrop 사용 / autoaugment / / momentum 0.9 / weight-decay 1e-4 / decay-rate0.1 /	DDP - 2개	200	81.02	36m 14s
cifar100	ResNet50	sgd / 256	+ lr 0.5 / image size 224 X 224	DDP - 2개	300	81.91	2h 22m 31s
cifar100	ResNet50	lion / 256	+ lr 5e-4 / weight-decay 1e-4 1e-3	DDP - 2개	400	77.25	3h 28m 27s