

## 24-1 vision팀 4월 정기 발표

---

Study Group vision  
2024/04/02

# Contents

---

- 3월 스터디 진행 overview
- 논문 스터디 내용
- 강의 스터디 내용
- 공모전 및 연구 주제 선정

# 3월 스터디 진행 overview

- 매주 화 오후 7시 사당역 오프라인 스터디 진행
- 매주 논문 발표 + 강의 발표로 진행
- 논문 큰 흐름 주제는 multi-modal 로 선정
- 강의는 유튜브 Deep Learning for Computer Vision(2020) 강의에서 선택하여 수강

논문 : Clip, Blip (문학준, 우다연)  
강의 : 12 : Recurrent Networks

논문 : DALL-E(이민서)  
강의 : 15 – Object Detection(인정교)

3/12

3/19

논문 : Oscar(김윤서)  
강의 : 13 – Attention(조재희)

3/23

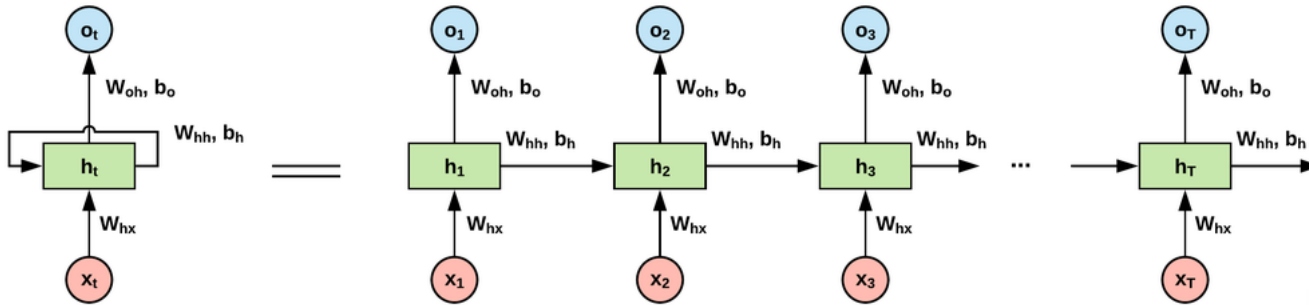
추가 주제 회의(줌 미팅)

3/26

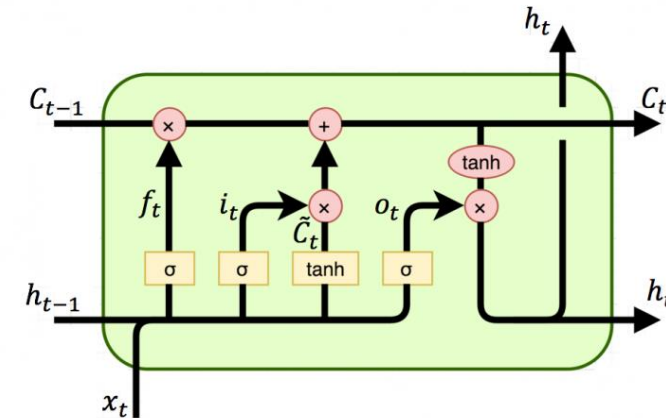
# 강의 스터디 내용

## 1. Lecutue 12 : Recurrent Networks

- Time series 데이터 처리를 위한 모델 구조
- 각 타임스텝 별 input용 가중치( $W_h$ ), 이전 hidden state용 가중치( $b_h$ )사용



- Gradient vanishing/exploding을 해결하기 위한 LSTM
- 3개의 게이트와 cell state로 구성



# 강의 스터디 내용

## 2. Lecuture 3 : Attention

- Seq2seq(Encoder&Decoder)
- > seq2seq with attention
- > transformers(\* self attention, positional Encoding, Multi-head cross attention)

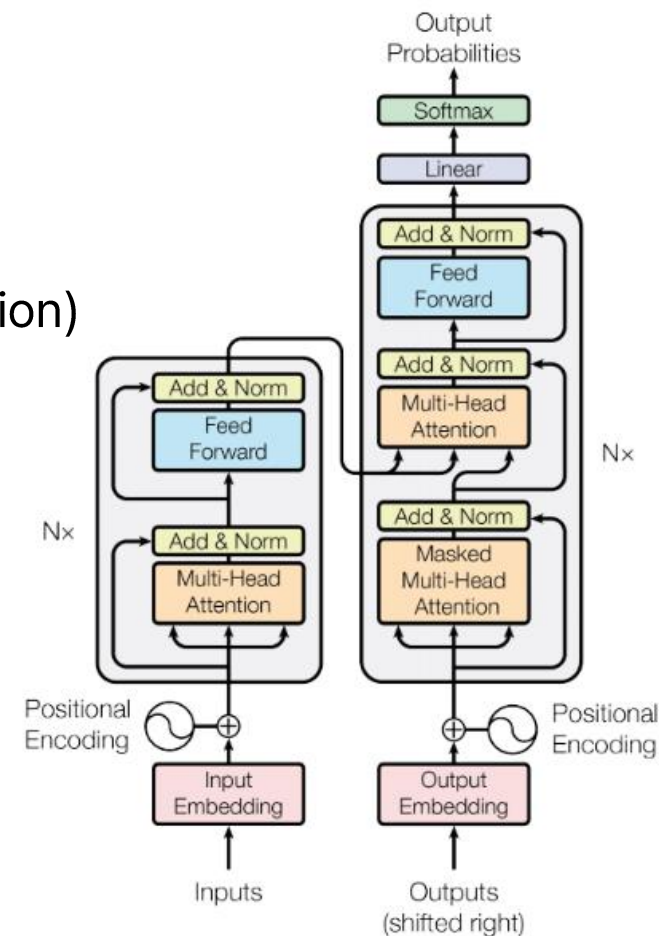
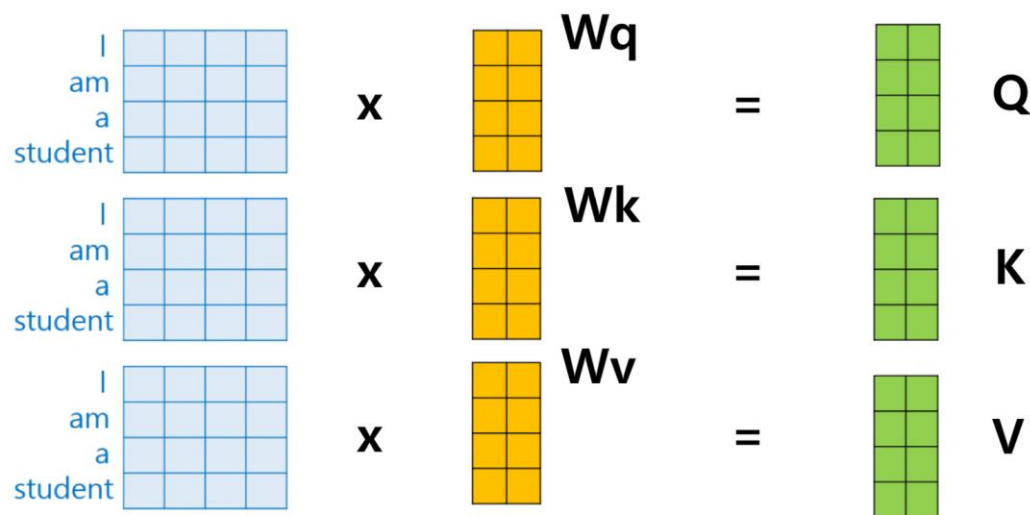
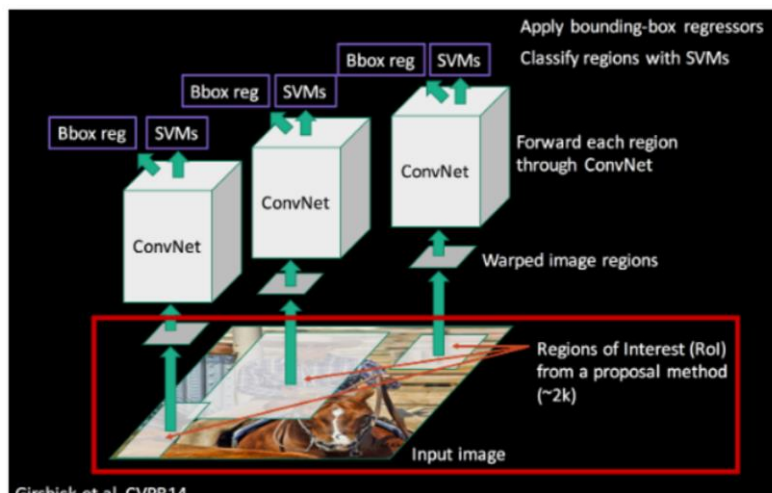


Figure 1: The Transformer - model architecture.

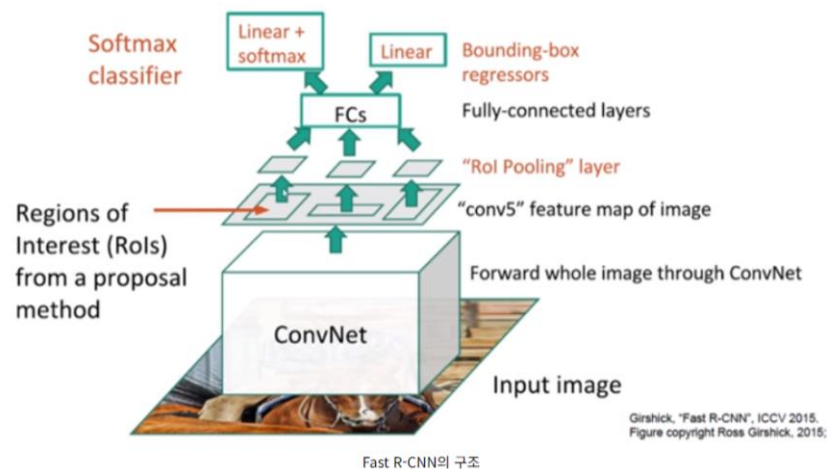
# 강의 스터디 내용

## 2. Lecutue 15 : Object Detection

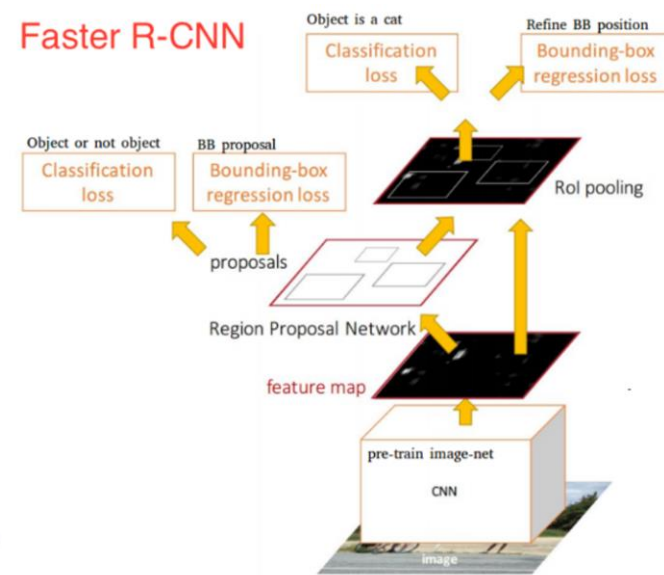
- 2 stage Detecting : input -> Region Proposal -> classification  
-> Bounding Box Regression



R-CNN



Fast R-CNN



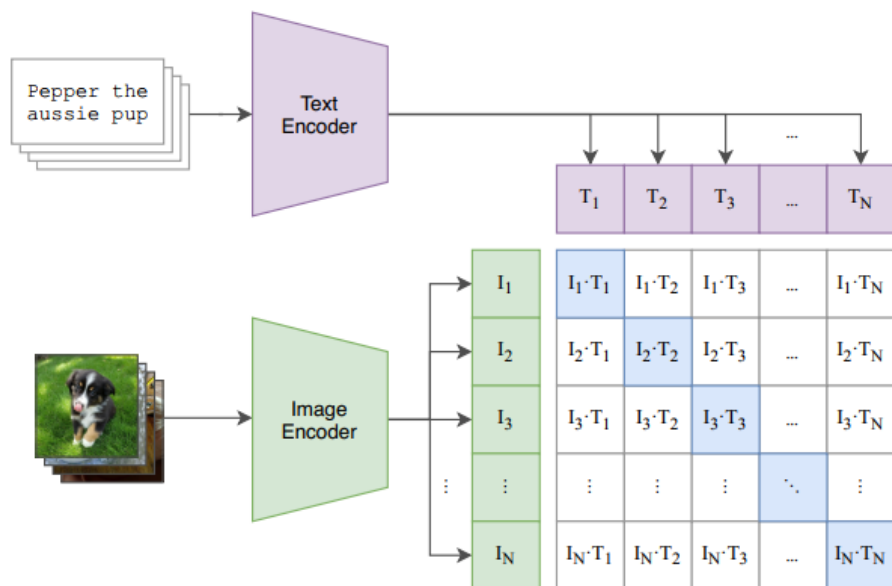
Faster R-CNN

# 논문 스터디 내용

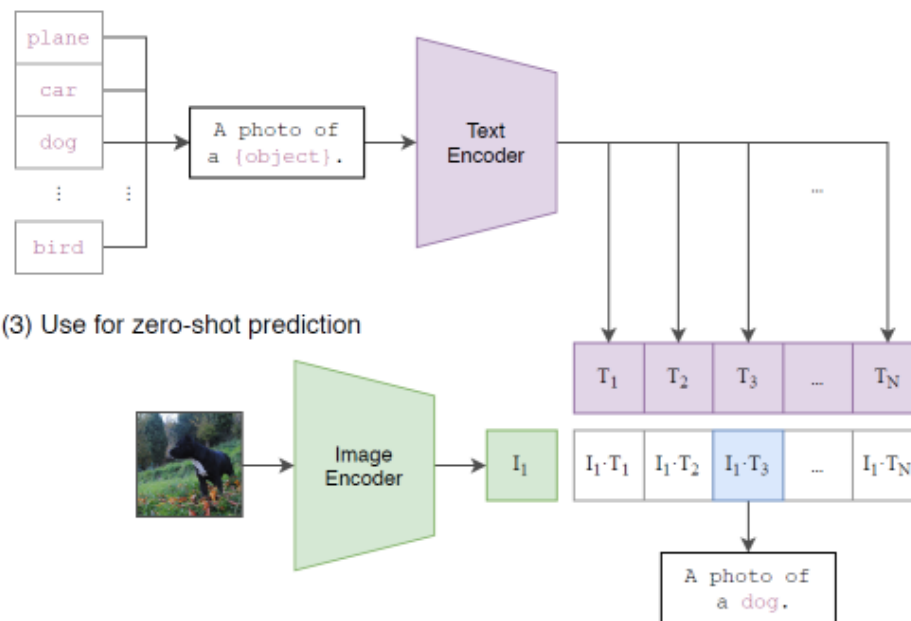
## 1. Clip(2021) / Blip(2022)

- Text embedding과 Image embedding을 연결하는 새로운 방법 제시 + Natural Language Supervision 이용
- Contrastive pre-training 적용

(1) Contrastive pre-training



(2) Create dataset classifier from label text

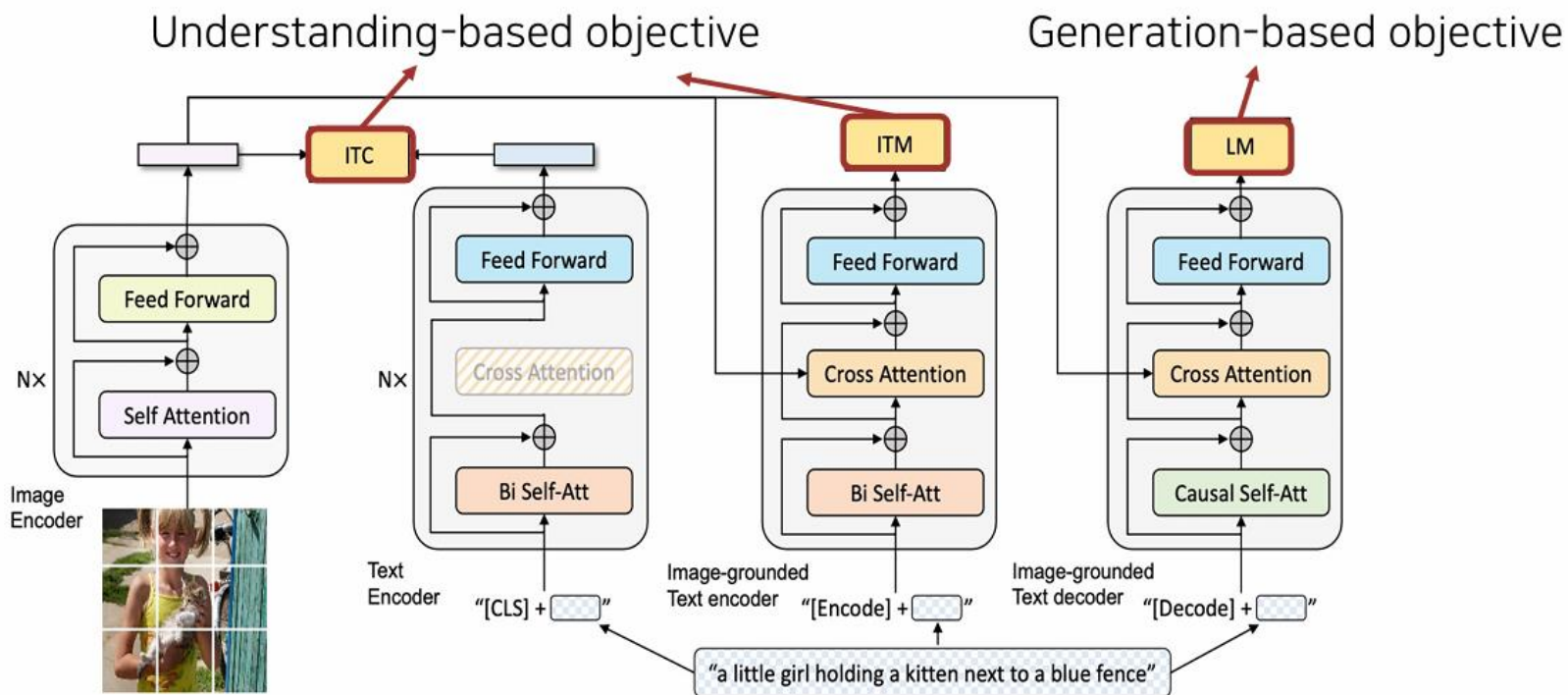


(3) Use for zero-shot prediction

# 논문 스터디 내용

## 1. Clip(2021) / Blip(2022)

- Encoder와 Decoder를 둘 다 사용하여 다양한 downstream task를 수행하도록 함
- Loss : Image-Text Contrastive loss + Image – Text Matching Loss + Language Modeling Loss

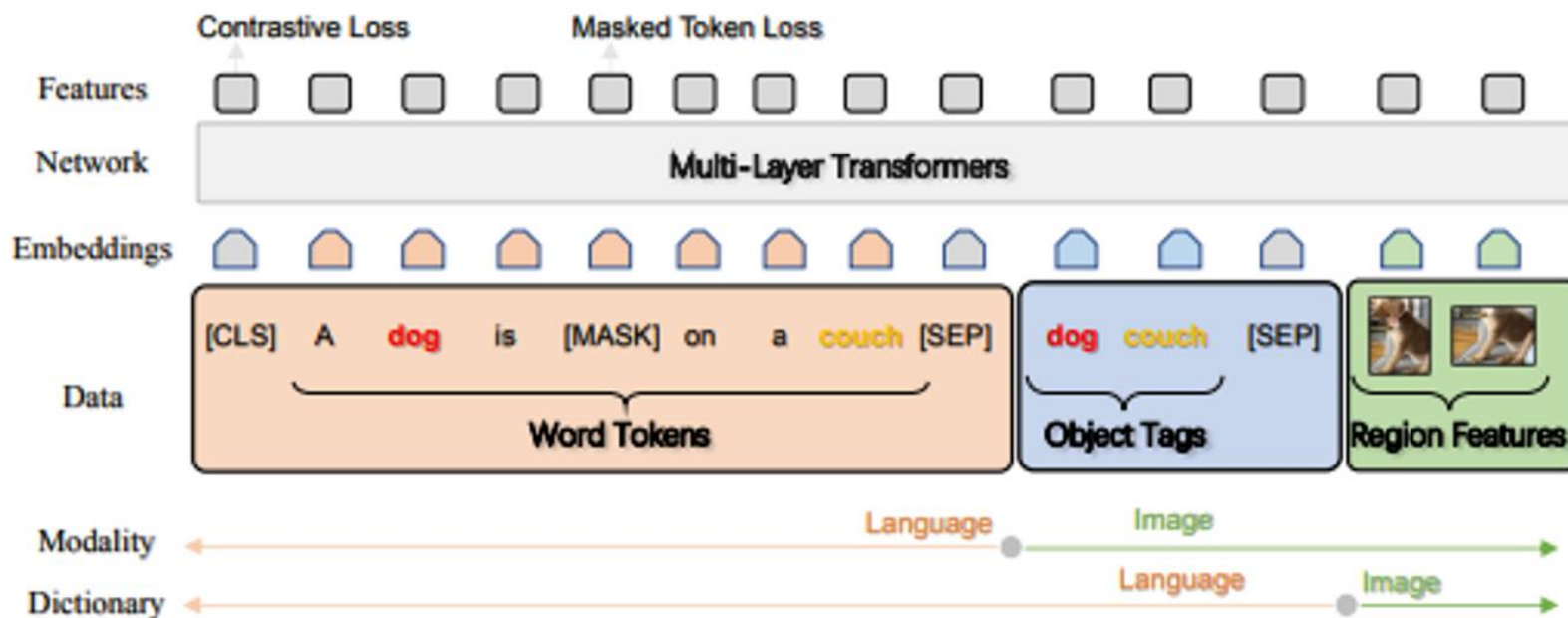




# 논문 스터디 내용

## 2. OSCAR(2020)

- 개별 인코더(BERT, Faster R-CNN)를 이용한 뒤 단일 트랜스포머에 적용하여 상호작용을 모델링하는 single stream 구조
- 기존 모델과 다르게 text embedding, image embedding과 함께 Tag embedding 까지 사용한다는 점이 큰 차이점



# 논문 스터디 내용

## 3. DALL-E(2021)


- Zero shot text-to-Image generation
- Transformer 모델을 단일 데이터 스트림 속에서 Auto-Regression 할 수 있도록 훈련
- 이미지 토큰 인풋의 메모리 한계를 극복하기 위해 two-stage 방법 채택(dVAE사용)

Text Prompt    an armchair in the shape of an avocado. . . .

AI Generated  
images



# 공모전 및 연구 주제 선정


 산업통상자원부 「창의융합형 공학인재 양성지원사업」

## 창의혁신 DNA HERO

### 캡스톤 디자인 연구팀 모집

**연구팀 구성** 팀장 1인 + 팀원 2~9인 + 지도교수 1인 + 산업체멘토 1인으로 이루어진 캡스톤 디자인 연구팀 **선착순 19팀**

**참여대상** 팀장 | 성균관대 전 계열 학부생  
 팀원 | 성균관대 전 계열 학부생 필수 (타 대학 학부생 참여 가능, 대학원생 참여 불가)

**지원규모** 팀당 240만원 내외

**연구기간** 2024. 4. ~ 2024. 9. (6개월)

**연구주제** 1. D.N.A(Data, Network, AI), 2. 바이오/디지털헬스케어, 3. 산업체 제안 주제, 4. 자유주제 (단, 본 사업 특화분야 DNA, 바이오/디지털헬스케어 분야 우선 선발)

산업분야	내용
DNA(Data, Network, AI) 기반 문제	Data, Network, AI 등 지능정보기술을 결합하여 새로운 가치를 창출하고자 하는 연구 문제
바이오/디지털헬스케어 관련 문제	바이오/디지털헬스케어(데이터 기반 의료 및 헬스케어 기기 개발 및 서비스)를 통해 인간의 지속가능한 건강한 삶에 기여하고자 하는 연구 문제
산업체 관심 기술 설계 문제	산업체(전문가)에서 제시하는 연구 주제 <sup>1)</sup> 혹은 학생이 산업체 현장 분석을 통해 발견한 스스로 발견한 산업체 연구 문제
기타 (자유주제)	창의적 역량을 시험하고 도전정신을 키울 수 있는 해결 방안 아이디어 제시

연구주제:  
 Transformers대체 기술인 **Mamba**를 활용한  
 style transfer 모델 연구



# 공모전 및 연구 주제 선정

## What is Mamba?

- 기존 transformers는 텍스트 길이가 길어짐에 따라 성능과 시간에 병목이 있음
  - ➔ 텍스트에 시간성을 부여하지 않고 압축함으로써 의미 함축을 잘 하지 못함
- 논문에서 제안하는 S6(Selective scan + S4) 구조를 사용하여 기존 S4 모델에 time-variance를 부여하고 Hardware-friendly 한 scan기법을 사용하여 gpu를 효율적으로 사용하도록 설계

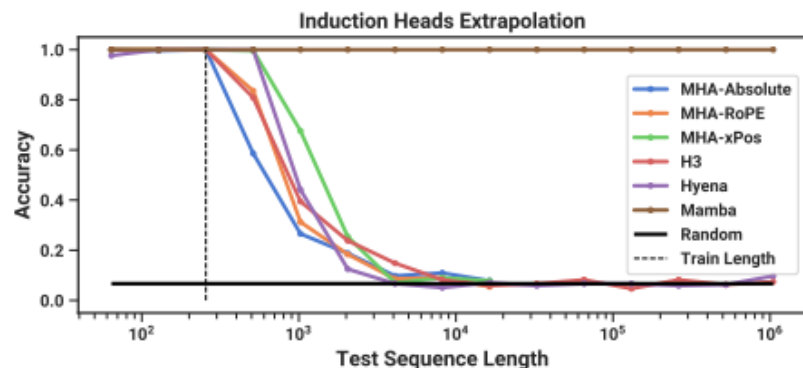


Table 2: (**Induction Heads.**) Models are trained on sequence length  $2^8 = 256$ , and tested on increasing sequence lengths of  $2^6 = 64$  up to  $2^{20} = 1048576$ . Full numbers in Table 11.

Feature	Transformer	Mamba
Architecture	Attention-based	SSM-based
Complexity	High	Lower
Inference Speed	$O(n)$	$O(1)$
Training Speed	$O(n^2)$	$O(n)$



TRAIN AND TEST