

Weakly Supervised Image Segmentation

김태미
2021.12.03.

Layer-Wise Relevance Propagation (LRP)

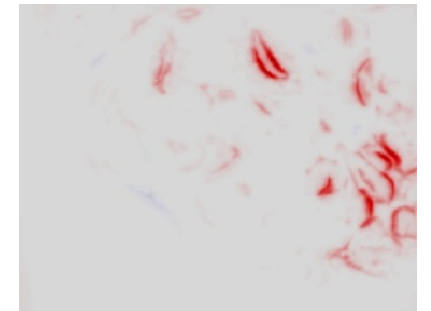
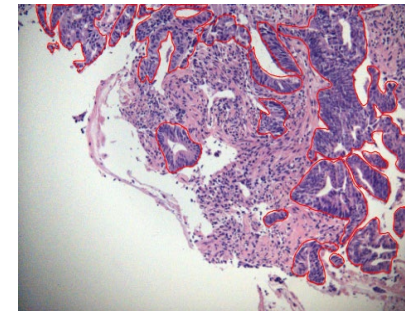
Malignant

- train: 6767 Images
- valid: 1422 Images

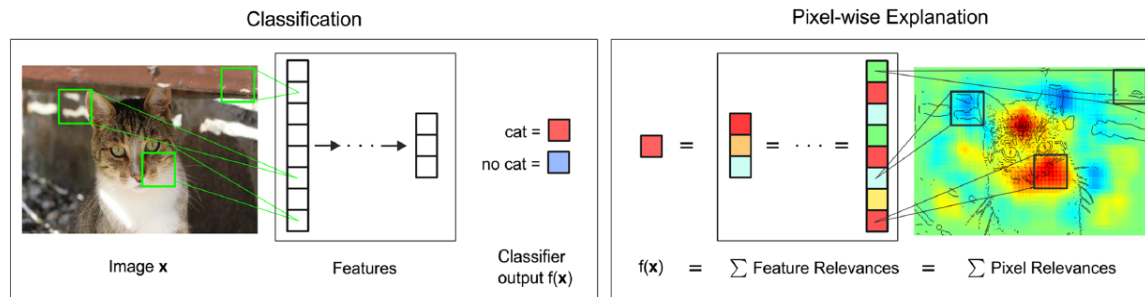
Normal

- train: 3813 Images
- valid: 1301 Images

- train: 10580 Images
- valid: 2723 Images



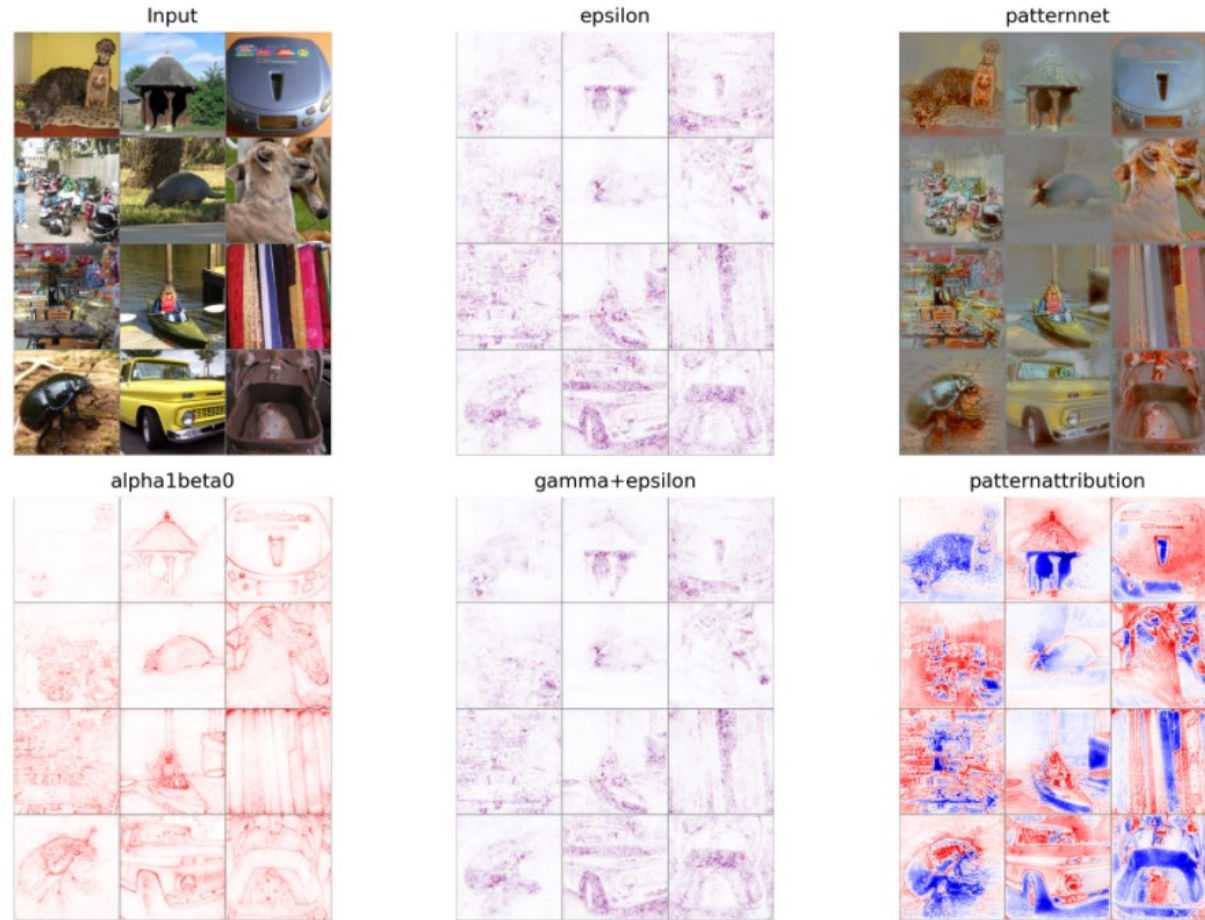
Relevance score
heat map
by LRP



Layer-Wise Relevance Propagation

LRP Rule

- epsilon rule
- gamma rule
- $\alpha=1$ $\beta=0$
- $\alpha=2$ $\beta=1$
- PatternAttribution
- PatternNet

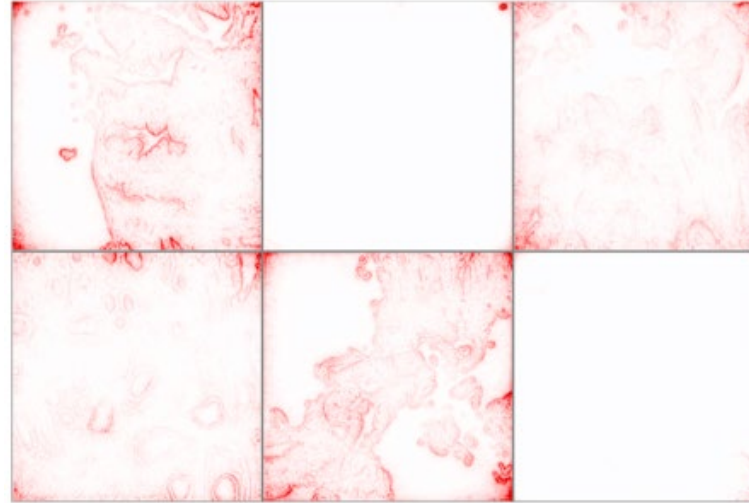


Layer-Wise Relevance Propagation

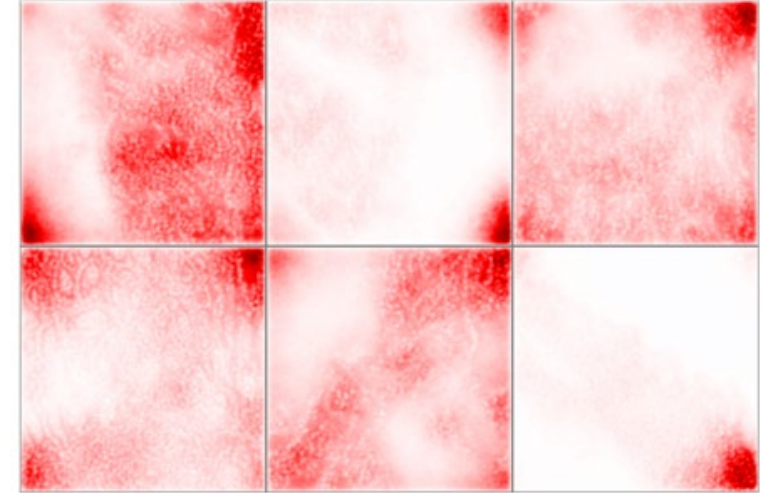
LRP Rule

- epsilon rule
- gamma rule
- $\alpha=1$ $\beta=0$
- $\alpha=2$ $\beta=1$
- PatternAttribution
- PatternNet

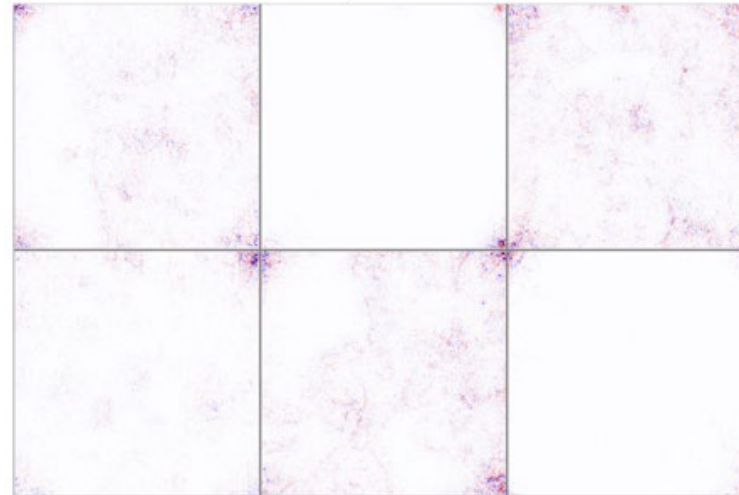
alpha1beta0



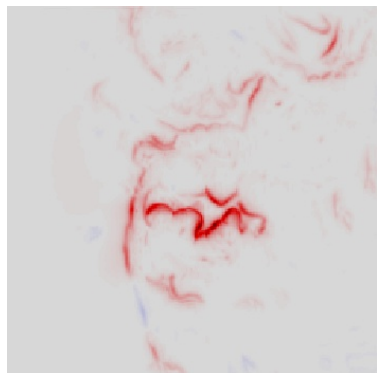
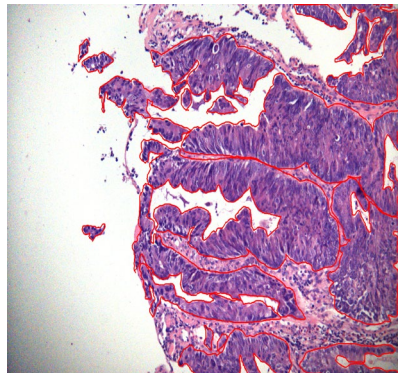
patternattribution



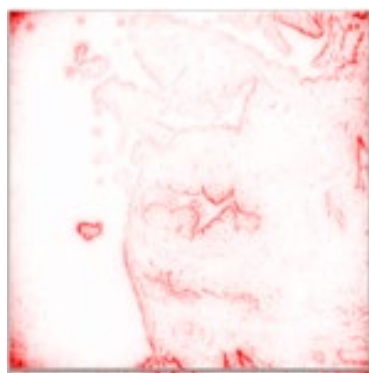
epsilon



Layer-Wise Relevance Propagation



기존 LRP
heatmap

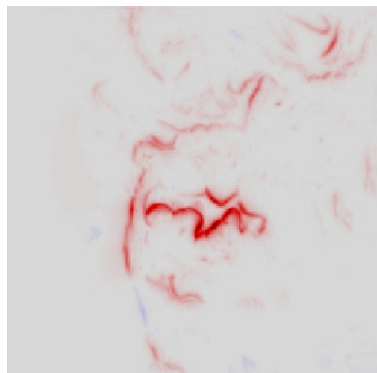
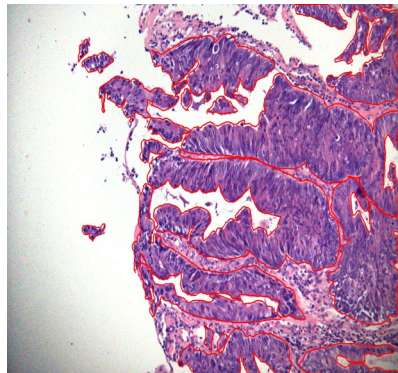


Rule 적용
heatmap

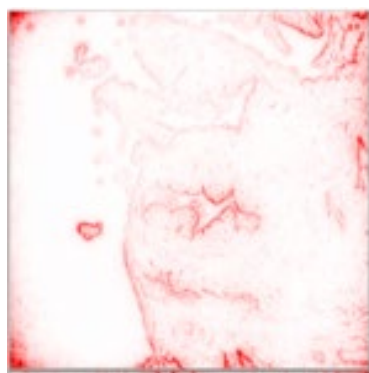
- lrp map이 malignant 부분과 상관 없이 전체 테두리 부분을 포착하는 경향이 많았다.
- 렌즈 그림자 부분을 포착하는 경우가 많았다.
- 기존 lrp heatmap 또한 경계선 위주로 포착하는 경향이 있다.

- ✓ 경계선을 이어서 pseudo label을 생성해보려 시도했지만, 경계선이 일관성이 없어서 보류.
- ✓ 기존 1000장에 비해 더 많이 10000장 가까이 train했지만 lrp 결과는 더 좋지않았다. -> malignant와 normal 이미지의 비율을 조정해봐야 할 것 같다.

Layer-Wise Relevance Propagation



기존 LRP
heatmap

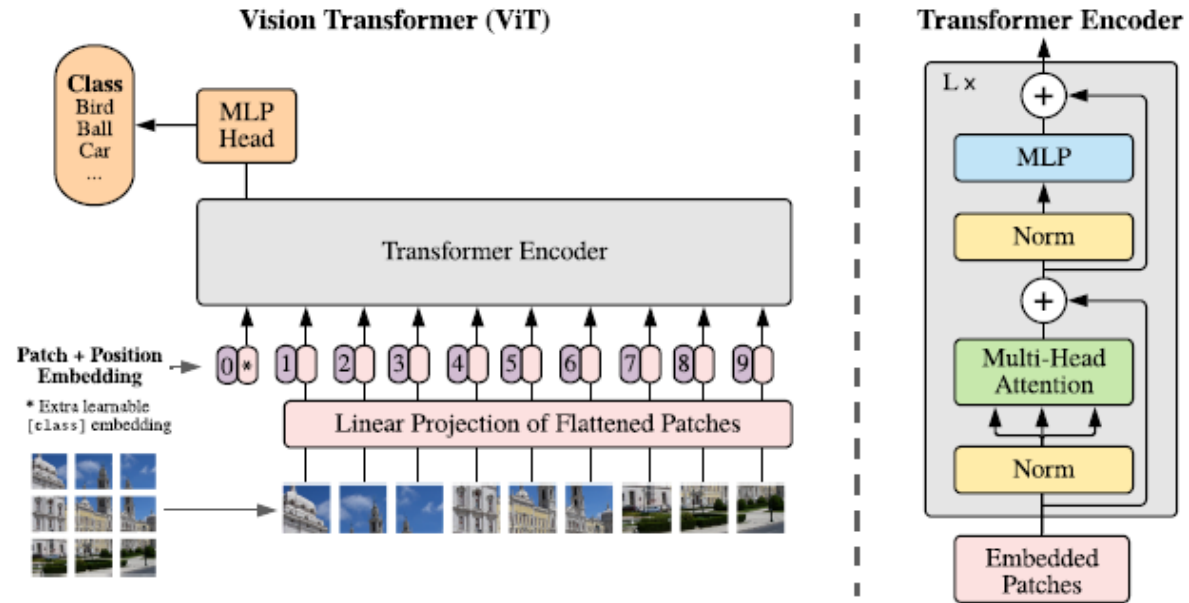


Rule 적용
heatmap

- lrp map이 malignant 부분과 상관 없이 전체 테두리 부분을 포착하는 경향이 많았다.
- 렌즈 그림자 부분을 포착하는 경우가 많았다.
- 기존 lrp heatmap 또한 경계선 위주로 포착하는 경향이 있다.

- ✓ 경계선을 이어서 pseudo label을 생성해보려 시도했지만, 경계선이 일관성이 없어서 보류.
- ✓ 기존 1000장에 비해 더 많이 10000장 가까이 train했지만 lrp 결과는 더 좋지않았다. -> malignant와 normal 이미지의 비율을 조정해봐야 할 것 같다.

Vision Transformer (ViT)



- Position Embedding
- Class Token
- Attention structure without RNN \rightarrow Transformer

Transformer

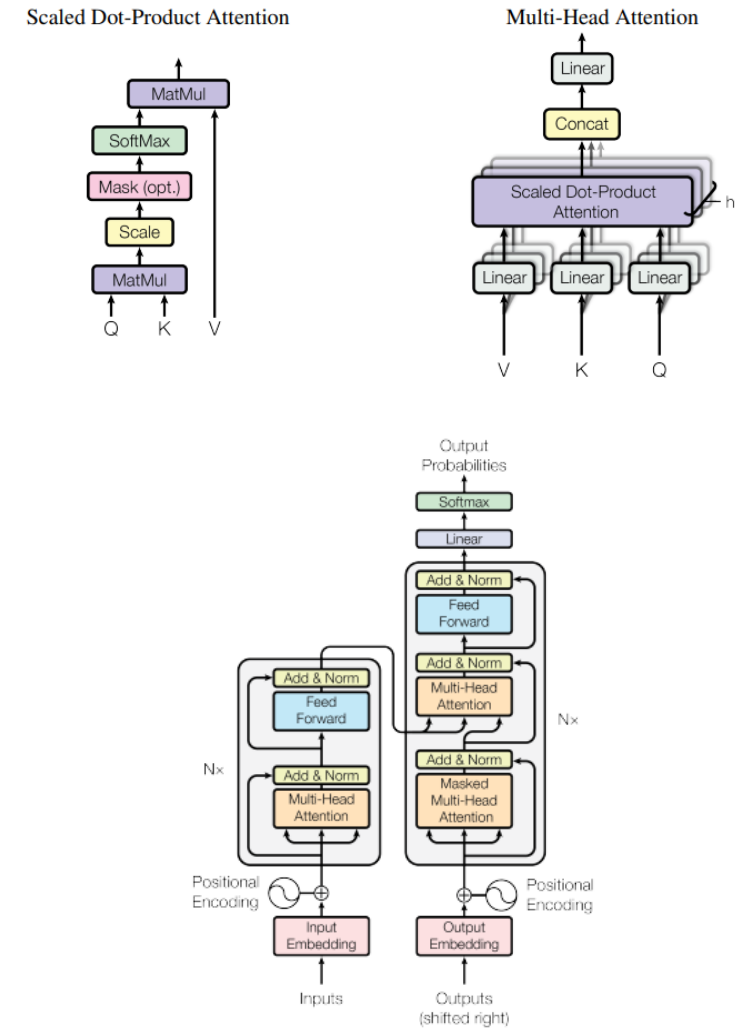
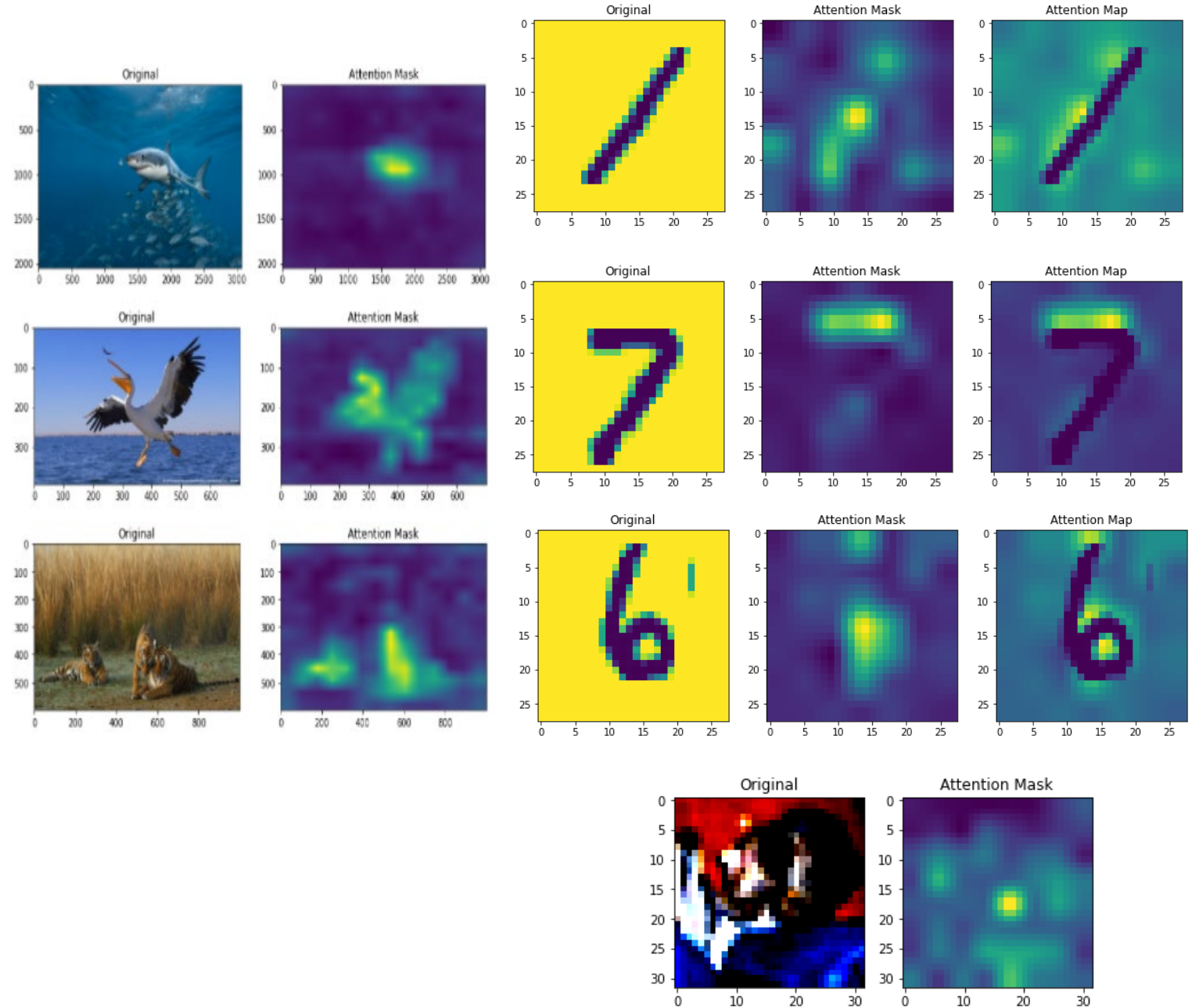


Figure 1: The Transformer - model architecture.

Vision Transformer (ViT)

Attention Map

- ViT는 CNN-based architecture 보다 inductive bias가 약해서 dataset이 적은 경우 성능이 떨어진다.
- 대용량 dataset으로 pretrained 된 경우에 CNN-based model의 성능을 보일 수 있다.
- ViT 구조 이해 및 구현이 목적, 디테일한 parameter 조정 및 성능 향상은 추후에 진행 예정.



Vision Transformer (ViT)

Voc 2012 데이터로 classification 진행

- Image size: 224 * 224
- Patch size: 16
- Head 수: 12
- Layer 수: 12
- Model dimension: 768
- Feedforward hidden dimension: 3072
- Optimizer: Adam → SGD 사용시 loss가 수렴되지 않았다.

50 epoch 실행 시 accuracy가 0.54 정도에서 더 이상 줄어들지 않고 증가하였다.

→ 낮은 classification accuracy로 인해 attention map이 제대로 뽑히지 않는다.

→ Pretrained된 ViT 모델을 이용하면
정확도가 훨씬 높다.

Vision Transformer (ViT)

Seegene Colon 데이터로 classification 진행

- Image size: 224 * 224
 - Patch size: 16
 - Head 수: 12
 - Layer 수: 12
 - Model dimension: 768
 - Feedforward hidden dimension: 3072
 - Optimizer: SGD/Adam
- Imagenet 으로 pretrained된 ViT 모델을 이용하였으나, loss가 줄어들지 않았다.
 - 현미경 이미지이므로 imagenet pretrained weight이 효과가 없다.
 - ViT는 대용량의 데이터로 훈련 시 accuracy 보장

...Thank you...