

실험 세팅

➤ 데이터 구성

- 360명의 환자 데이터 (가톨릭 전체 데이터)
- Neoplastic : 53명(174장), Non-neoplastic : 307명(873장)

➤ Train & Test Set

- 1,047장의 데이터셋
- Train Set : 288명(838장)
- Test Set : 72명(209장)

➤ Input Image

- Outline, Inline, Polyp, 담낭벽+Polyp 이미지

➤ 모델 구성

- 갑상선에서 사용한 모델을 담낭 용종 연구에 맞게 조정한 모델
- 데이터 불균형 문제를 해소하기 위한 Focal loss function 적용

➤ 30번 반복 실험

Analysis

1. 입력 이미지 검토

- 전처리 후 입력 이미지 검토
- 전처리 과정 문제없음

2. Activation Map

- Activation Map을 통한 중요 판단 부분 확인

3. Data Augmentation

- Data augmentation을 통해 데이터 부족으로 인한 문제인지 확인

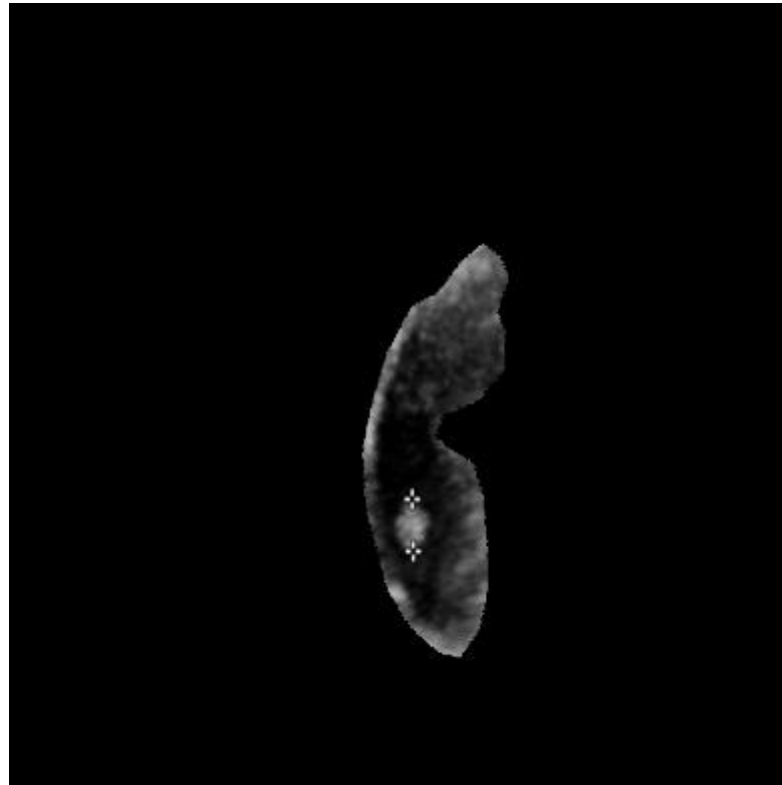
4. 각 입력 이미지에 대한 실험

- 각 입력 이미지별 실험 성능 확인

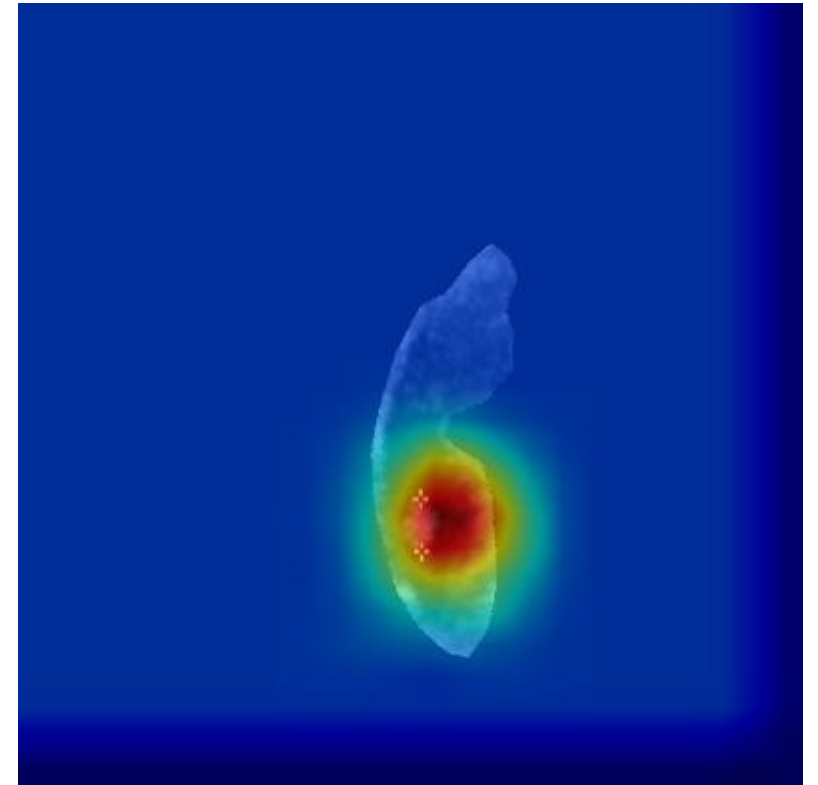
Class Activation Map

- 입력 이미지에 대한 Class Activation Map(CAM)
- 입력 이미지에 대한 중요 판단 부분 시각화

Input Image



CAM Result



Best Case

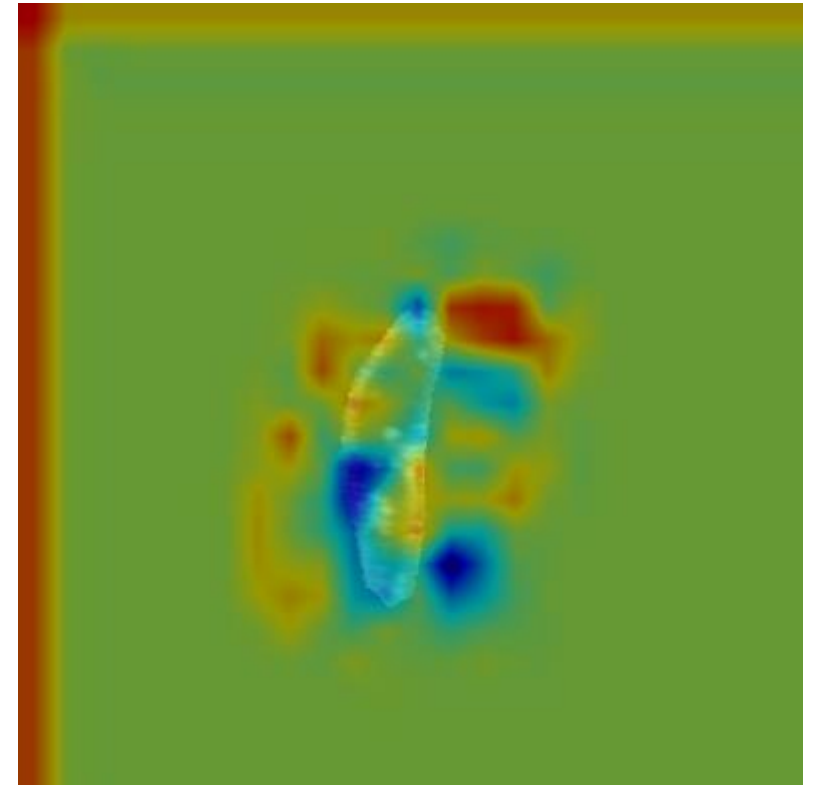
Class Activation Map

- 입력 이미지에 대한 Class Activation Map(CAM)
- 입력 이미지에 대한 중요 판단 부분 시각화

Input Image



CAM Result



Worst Case

Experiment Results

➤ Class Activation Map

- Best case의 경우 오차 범위가 작은 이미지에 대한 activation map
- Activation map이 polyp을 중요 판단 부분으로 인식
- Worst case의 경우 오차 범위가 가장 큰 이미지에 대한 activation map
- Worst의 경우에도 일부 이미지에 대해서는 polyp 근처를 중요 판단 부분으로 인식
- Multi polyp의 경우 polyp과 polyp 사이를 중요 판단 부분으로 인식
- 또는 완전히 다른 부분을 중요 판단 부분으로 인식

Top-3 Vote Method Classification Performance Table

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Baseline Model based on paired t-test at 5% significance level).

Model	Metrics	AUC Mean	95% CI		p-value
			Lower	Upper	
Baseline	AUC	0.597	0.561	0.634	
	Accuracy	0.619	0.521	0.717	
	Sensitivity	0.370	0.220	0.520	
	Specificity	0.664	0.522	0.806	
Data Augmentation	AUC	0.616	0.582	0.651	0.548
	Accuracy	0.536	0.430	0.643	0.379
	Sensitivity	0.488	0.342	0.634	0.384
	Specificity	0.545	0.393	0.696	0.376
All Data	AUC	0.798	0.780	0.816	0.001 *△
	Accuracy	0.524	0.449	0.599	0.253
	Sensitivity	0.777	0.678	0.877	0.017 △
	Specificity	0.422	0.283	0.561	0.103

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001
Data Augmentation : Random Horizontal Flip, All Data : 서울대 데이터 & 가톨릭 데이터

Experiment Results

➤ Data Augmentation

- 가톨릭 데이터에 Random Horizontal Flip 기법을 적용하여 실험 진행
- 기존 실험 결과에 비해 약간의 성능 향상
- 서울대 데이터와 가톨릭 데이터 모두 사용하여 실험 진행
- 서울대 데이터만 사용하였을 경우 Top-3 기준 AUC는 0.925
- 모든 데이터를 사용한 결과 큰 폭으로 성능 하락
- Top-3 기준 Accuracy와 Specificity의 성능이 대폭 하락
- 데이터 부족으로 인한 성능 저하는 원인이 아님
- 서울대 데이터와 가톨릭 데이터 차이점 분석 필요

Top-3 Vote Method Classification Performance Table

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Polyp Image based on paired t-test at 5% significance level).

Input Image	Metrics	AUC Mean	95% CI		p-value
			Lower	Upper	
Outline	AUC	0.526	0.491	0.560	0.190
	Accuracy	0.839	0.824	0.855	0.440
	Sensitivity	0.006	0.000	0.018	0.205
	Specificity	0.990	0.969	1.000	0.209
Inline	AUC	0.566	0.534	0.598	0.866
	Accuracy	0.825	0.801	0.848	0.175
	Sensitivity	0.048	0.007	0.090	0.114
	Specificity	0.964	0.931	0.998	0.139
Polyp	AUC	0.571	0.535	0.607	
	Accuracy	0.847	0.847	0.847	
	Sensitivity	0.000	0.000	0.000	
	Specificity	1.000	1.000	1.000	
담낭벽 + Polyp	AUC	0.543	0.510	0.575	0.369
	Accuracy	0.846	0.844	0.848	0.435
	Sensitivity	0.000	0.000	0.000	-
	Specificity	0.999	0.997	1.000	0.205

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001

Experiment Results

➤ 각 이미지별 실험

- 각 입력 이미지별 실험 진행
- Top-3 기준 AUC 실험 결과는 모든 입력 이미지를 사용한 결과와 유사
- Specificity 결과가 높은 것을 보았을 때 Non-neoplastic에 대한 판단 능력은 문제없음
- Sensitivity 결과의 경우 Neoplastic에 대해 못 맞추고 있음
- Neoplastic 데이터에 대한 Activation Map 결과는 대부분 polyp을 중요 판단 부분으로 인식
- Neoplastic 데이터에 대한 분석 필요