

Multimodal Datasets Search



AutoML 연구실
석사과정 문아성
21.11.30

TO DO List

➤ Last week

- Multimodal Datasets search – Modality가 애매한 Dataset 제외
- 담낭용종 - 기존 모델의 낮은 성능에 대한 원인 분석
 - 기존 서울대 데이터 실험 AUC: 0.905, Accuracy: 0.832
 - 신규 서울대 데이터 실험 AUC: 0.621, Accuracy: 0.510
 - 데이터 문제, 내부 모델 문제
- ~~심미성 후속 연구 제안서~~

➤ This week

- Multimodal Datasets search – 추가 Multimodal Dataset 수집
- 담낭용종 – 데이터 전처리 후
 - 기존 서울대 데이터 실험 AUC: 0.905, Accuracy: 0.832
 - 신규 서울대 데이터 실험 AUC: 0.694, Accuracy: 0.705

Multimodal Datasets Search

Multimodal Datasets List (1/2)

Dataset	Modality	Pattern	Features	Annotation	Domain
CUB-200-2011	Image, binary attributes	11,788	28 (<i>image per</i>)	200	Birds
UPMC Food-101	Image, text	90,704	93,409	101	Food
MuMu Dataset	Audio, image	147,295	31,471	20	Music
Oxford Flower-102	Image, tag	8,189	102	102	Flower
Web Queries	Image, text	71,478	353	2	Web Query
PASCAL VOC'07	Image, tag	10,000	804	20	Object
MIR Flickr	Image, tag	25,000	457	38	Object
CrisisMMD	Image, text	12,743	11,404	2	Crisis
GTZAN	Image, MIR features	1,000	888	10	Music
Ballroom	Image, MIR features	698	888	8	Music

Multimodal Datasets Search

Multimodal Datasets List (2/2)

Dataset	Modality	Pattern	Features	Annotation	Domain
Hateful Memes	Image, text	10,000	10,000	2	Hatred
RVL-CDIP	Image, text	400,000	-	16	Document
Tobacco3482	Image, text	3,482	-	10	Document
NUS-WIDE	Image, tag	269,648	5,018	81	Object
Wikipedia	Image, text	2,866	2,866	10	Object

The Caltech-UCSD Birds-200-2011 Dataset (*CUB-200-2011 Dataset*)

- Bird species classification
- Modality: Image, Binary Attributes



Examples of images in the UCB-200-2011 Dataset

- Dataset Specification
 - Images: 11,788
 - Bird species: 200
 - Bird attributes: 28
 - Part Locations: 15

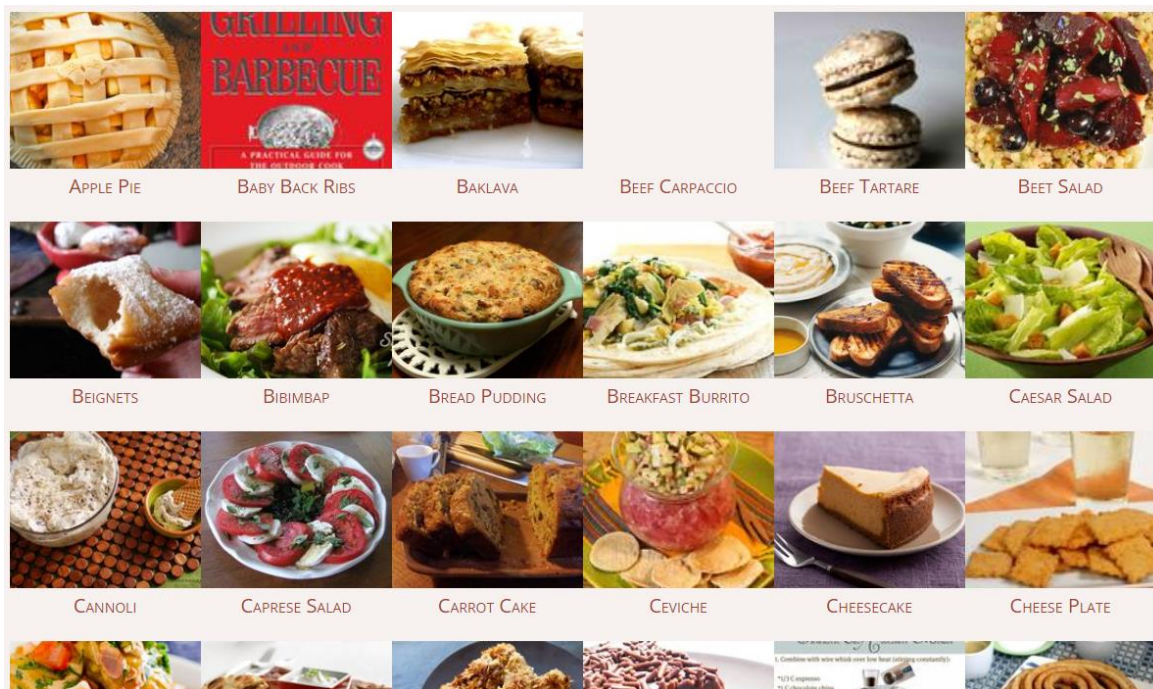
Part	Attributes	Part	Attributes	Part	Attributes
Beak	HasBillShape, HasBillColor, HasBillLength	Back	HasBackColor, HasBackPattern	Breast	HasBreastPattern, HasBreastColor
Belly	HasBellyPattern, HasBellyColor	Fore-head	HasForeheadColor	Bird (all parts)	HasSize, HasShape
Throat	HasThroatColor	Nape	HasNapeColor	Head	HasHeadPattern
Crown	HasCrownColor	Eye	HasEyeColor	Leg	HasLegColor
Tail	HasUpperTailColor, HasUnderTailColor, HasTailPattern, HasTailShape	Wing	HasWingPattern, HasWingColor, HasWingShape	Body	HasUnderpartsColor, HasUpperPartsColor, HasPrimaryColor

Attribute Part Associations

Multimodal Datasets Search

UPMC Food-101 Dataset

- Food categories classification
- Modality: Image, text(recipe)



Examples of images in the UPMC Food-101 Dataset

- Dataset Specification
 - Number of images: 90,704 (train: 67,988 test: 22,716)
 - Number of text: 93,409
 - annotation: 101



Text(recipe) data sample: Apple Pie

Multimodal Datasets Search

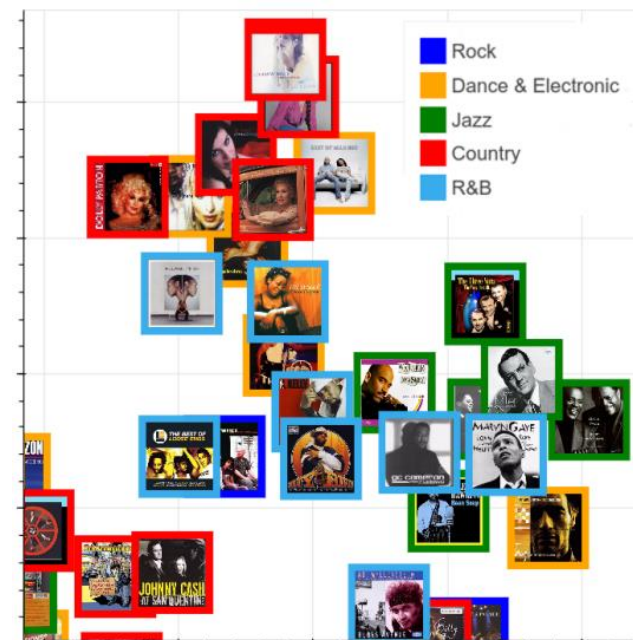
MuMu Dataset

- *Music genre classification*
- *Modality: Audio, image(album cover)*

Genre	% of albums	Genre	% of albums
Pop	84.38	Tributes	0.10
Rock	55.29	Harmonica Blues	0.10
Alternative Rock	27.69	Concertos	0.10
World Music	19.31	Bass	0.06
Jazz	14.73	European Jazz	0.06
Dance & Electronic	12.23	Piano Blues	0.06
Metal	11.50	Norway	0.06
Indie & Lo-Fi	10.45	Slide Guitar	0.06
R&B	10.10	East Coast Blues	0.06
Folk	9.69	Girl Groups	0.06

Top-10 most and least represented genres

- *Dataset Specification*
 - Number of audio: 147,295
 - Number of images: 31,471
 - annotation: 20



Particular of the t-SNE of randomly selected image vectors from five of the most frequent genres.

Oxford Flower-102 Dataset

- Flower categories classification
- Modality: image, text(caption)



Examples of images in the Oxford Flower-102 Dataset

- Dataset Specification
 - Number of image: 8,189
 - Number of text: 102
 - annotation: 102

Caption	Image
this flower has white petals and a yellow stamen	
the center is yellow surrounded by wavy dark purple petals	
this flower has lots of small round pink petals	

Examples of generated images from text descriptions.

Multimodal Datasets Search

Web Queries Dataset

- Web image relevance *classification*
- *Modality: image, text(query)*

- *Dataset Specification*
 - Number of image: 71,478
 - Number of text: 353
 - annotation: 2 (*relevant / irrelevant*)

Queries	Images			
"arc de triomphe"				
"eiffel tower"				
"orsay museum"				

Examples of images in the Web Queries Dataset

Query: "arc de triomphe"



relevant to the query



irrelevant to the query

Examples of web images relevance to the query

Multimodal Datasets Search

PASCAL VOC'07 Dataset

- *Object categories classification*
- *Modality: image, tag*

- *Dataset Specification*
 - Number of image: 10,000
 - Number of tag: 804
 - annotation: 20



Examples of images in the PASCAL VOC'07 Dataset



Tags: india
Labels: cow

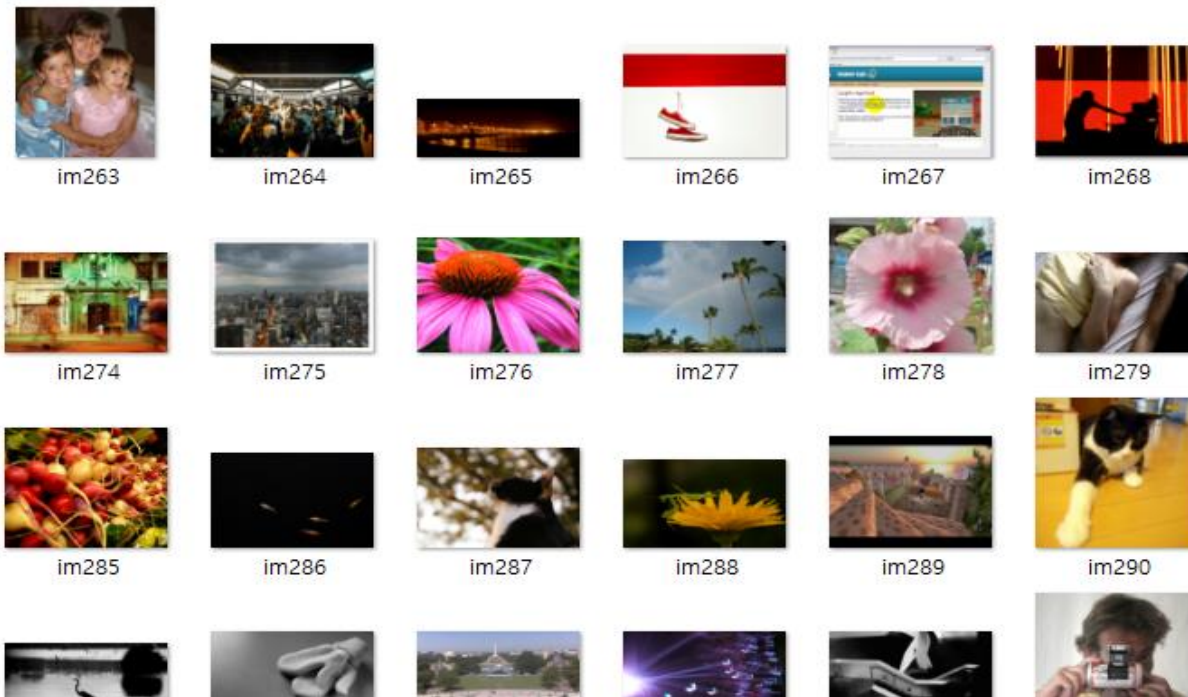


Tags: aviation, airplane, airport
Labels: aeroplane

Example images VOC'07 dataset
with their associated tags and class labels

MIR Flickr Dataset

- *Object categories classification*
- *Modality: image, tag*



Examples of images in the MIR Flickr Dataset

- *Dataset Specification*
 - Number of image: 25,000
 - Number of tag: 457
 - annotation: 38



Tags: desert,nature,landscape,sky
Labels: clouds, plant life, sky, tree



Tags: rose, pink
Labels: flower, plant life

Example images MIR Flickr dataset
with their associated tags and class labels

CrisisMMD Dataset

- Crisis relevance classification
- Modality: image, text

- Dataset Specification
 - Number of image: 12,743
 - Number of text: 11,404
 - annotation: 2

Images



Text

Crisis Name	Images		Text Messages	
	Informative	Not Informative	Informative	Not Informative
Hurricane Irma	2018	766	1836	678
Hurricane Harvey	2258	906	2082	800
Hurricane Maria	1813	1295	1594	1139
California wildfires	923	282	873	261
Mexico earthquake	806	315	732	285
Iraq-Iran earthquake	398	102	330	83
Sri Lanka floods	229	632	184	527
Total	8445	4298	7631	3773

Annotation

Not informative	Informative	Informative	Informative
-----------------	-------------	-------------	-------------

Sample images along with their text and annotations

Crisis Name	Images		Text Messages	
	Informative	Not Informative	Informative	Not Informative
Hurricane Irma	2018	766	1836	678
Hurricane Harvey	2258	906	2082	800
Hurricane Maria	1813	1295	1594	1139
California wildfires	923	282	873	261
Mexico earthquake	806	315	732	285
Iraq-Iran earthquake	398	102	330	83
Sri Lanka floods	229	632	184	527
Total	8445	4298	7631	3773

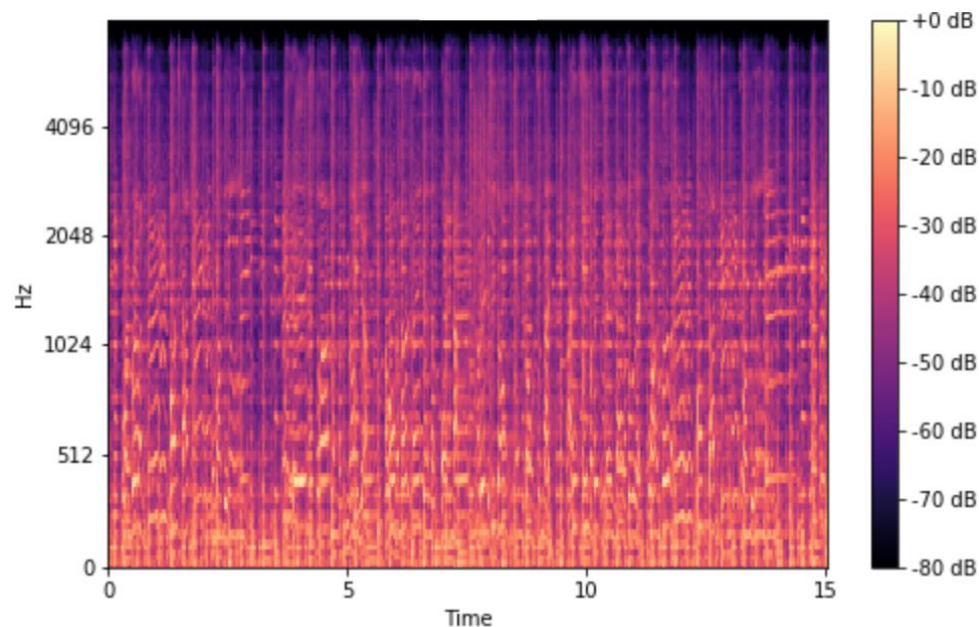
The number of images and text messages in the dataset

GTZAN / Ballroom Dataset

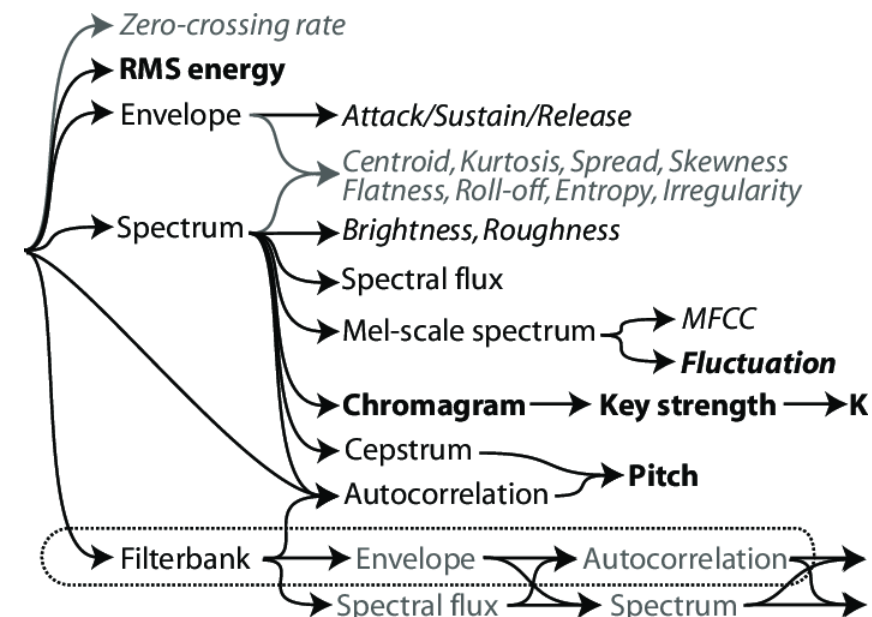
- *Music genre classification*
- *Modality: image, acoustic features*

- *Dataset Specification*

- Number of image: *GTZAN- 1,000, Ballroom- 698*
- Number of *acoustic features* : 888
- annotation: *GTZAN- 10, Ballroom- 8*



Mel Spectrogram Image



Music acoustic features that can be extracted with MIRToolbox

Facebook Hateful Meme Dataset

- *Hateful Meme classification*
- *Modality: image, text*
- *Dataset Specification*
 - Number of image: 10,000
 - Number of *text* : 10,000
 - annotation: 2



Examples of images in the Hateful Meme Dataset

To automatically classify memes as hateful or not.

Multimodal Datasets Search

RVL-CIP Dataset

- Document image classification
- Modality: image, text
- Dataset Specification
 - Number of image: 400,000
 - Text embedding vector
 - annotation: 16

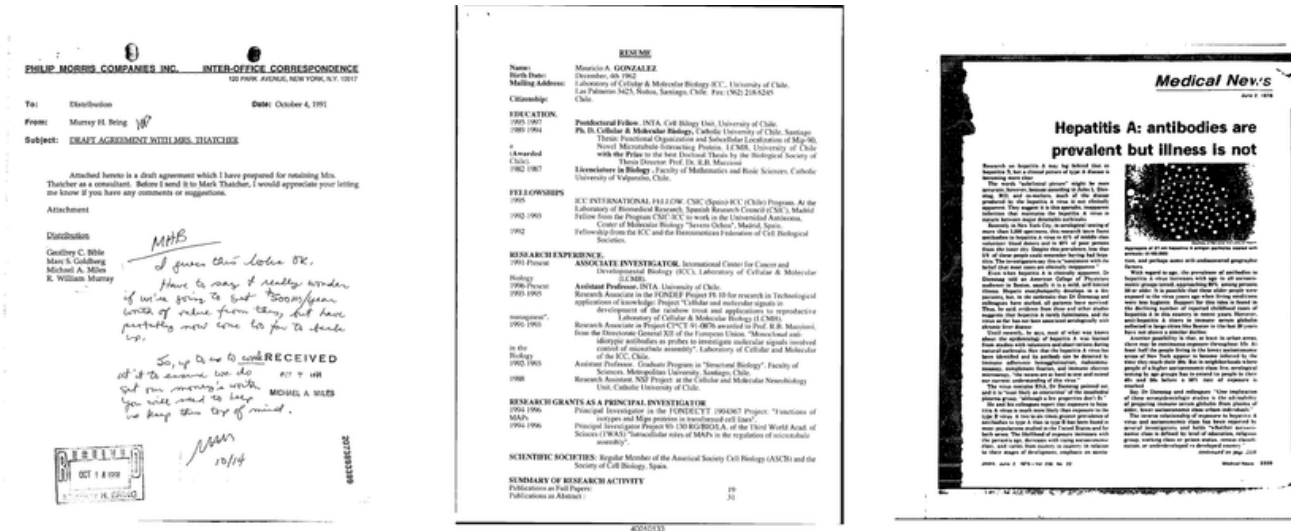


Examples of images in the RVL-CIP Dataset

Multimodal Datasets Search

Tobacco3482 Dataset

- Document image classification
- Modality: image, text
- Dataset Specification
 - Number of image: 3482
 - Text embedding vector
 - annotation: 10



Examples of images in the Tobacco3482 Dataset

NUS-WIDE Dataset

- *Object categories classification*
- *Modality: image, tag*
- *Dataset Specification*
 - Number of image: 269,648
 - Number of text: 5,018
 - annotation: 81



(a) sunset, sea



(b) sunset, sea



(c) desert, tree



(d) sunset, tree



(e) sunset, tree



(f) mountain, desert

Examples of images in the NUS-WIDE Dataset

Wikipedia Dataset

- *Object categories classification*
- *Modality: image, text*
- *Dataset Specification*
 - Number of image: 2,866
 - Number of text: 2,866
 - annotation: 10



All nuthatches nest in cavities; except for the two species of rock nuthatches, all use tree holes, making a simple cup lined with soft materials on which to rest eggs. In some species the lining consists of small woody objects such as bark flakes and seed husks, while in others it includes the moss, grass, hair and feathers typical of passerine birds...



Bramall Hall, a Grade I listed building, is a 14th-century black and white timber framed Tudor manor house, located between Cheadle Hulme and Bramhall. Described by Stockport Metropolitan Borough Council (SMBC) as "the most prestigious and historically significant building in the Conservation Area"...

Examples of images in the Wikipedia Dataset

Reference

- CUB-200-2011 dataset

*Wah, Catherine, et al. "The caltech-ucsd birds-200-2011 dataset." (2011).
<http://www.vision.caltech.edu/visipedia/CUB-200-2011.html>*

- UPMC Food-101 dataset

*Wang, Xin, et al. "Recipe recognition with large multimodal food dataset."
2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2015.
<http://visiir.lip6.fr/>*

- MuMu Dataset

*Oramas, Sergio, et al. "Multi-label music genre classification from audio, text, and images using deep features."
arXiv preprint arXiv:1707.04916 (2017).
<https://www.upf.edu/web/mtg/mumu>*

- Oxford Flower-102 Dataset

*Nilsback, Maria-Elena, and Andrew Zisserman. "Automated flower classification over a large number of classes."
2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing. IEEE, 2008.
<https://www.robots.ox.ac.uk/~vgg/data/flowers/102/>*

- Web Queries Dataset

<http://lear.inrialpes.fr/~krapac/webqueries/webqueries.html>

Reference

- PASCAL VOC'07 Dataset

M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007.
<http://host.robots.ox.ac.uk/pascal/VOC/voc2007/>

- MIR Flickr Dataset

Huiskes, Mark J., and Michael S. Lew. "The mir flickr retrieval evaluation."
Proceedings of the 1st ACM international conference on Multimedia information retrieval. 2008.
<https://press.liacs.nl/mirflickr/>

- CrisisMMD Dataset

Zou, Zhiqiang, et al. "Disaster Image Classification by Fusing Multimodal Social Media Data." ISPRS International Journal of Geo-Information 10.10 (2021): 636.
<https://crisisnlp.qcri.org/crisismmd>

- GTZAN Dataset

Sturm, Bob L. "The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use." arXiv preprint arXiv:1306.1461 (2013).
<http://marsyas.info/index.html>

- Web Queries Dataset

Marchand, Ugo, and Geoffroy Peeters. "The extended ballroom dataset." (2016).
<http://mtg.upf.edu/ismir2004/contest/tempoContest/node5.html>

Reference

- Hateful Memes Dataset

Kiela, Douwe, et al. "The hateful memes challenge: Detecting hate speech in multimodal memes." arXiv preprint arXiv:2005.04790 (2020).
<https://ai.facebook.com/blog/hateful-memes-challenge-and-data-set/>

- RVL-CIP Dataset

Harley, Adam W., Alex Ufkes, and Konstantinos G. Derpanis. "Evaluation of deep convolutional nets for document image classification and retrieval." 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015.
<https://www.cs.cmu.edu/~aharley/rvl-cdip/>

- Tobacco3482 Dataset

Kumar, Jayant, Peng Ye, and David Doermann. "Structural similarity for document image classification and retrieval." Pattern Recognition Letters 43 (2014): 119-126.

- NUS-WIDE Dataset

Chua, Tat-Seng, et al. "Nus-wide: a real-world web image database from national university of singapore." Proceedings of the ACM international conference on image and video retrieval. 2009.
<https://lms.comp.nus.edu.sg/wp-content/uploads/2019/research/nuswide/NUS-WIDE.html>

- Wikipedia Dataset

Rasiwasia, Nikhil, et al. "A new approach to cross-modal multimedia retrieval." Proceedings of the 18th ACM international conference on Multimedia. 2010.
<http://www.svcl.ucsd.edu/projects/crossmodal/>

➤ Last week

- 기존 모델의 낮은 성능에 대한 원인 분석
기존 서울대 데이터 실험 AUC: 0.905, Accuracy: 0.832
신규 서울대 데이터 실험 AUC: 0.621, Accuracy: 0.510
- 데이터 문제, 내부-모델-문제

➤ This week

- 데이터 검토 및 전처리
- 서울대 신규 데이터 TOP-3 실험

- 기존 서울대 병원 데이터와
신규 서울대 병원 데이터와의 이미지 크기 차이

대체로 신규 서울대병원 데이터의 이미지 크기가 더 큼
기존 데이터와 이미지 비율(높이, 너비)이 다름

- 정보 손실을 최소화하기 위한 이미지 사이즈 설정 필요

(높이, 너비)	이미지 수
(300, 510)	1
(302, 516)	2
(308, 507)	1
(315, 456)	1
(320, 516)	1
(327, 516)	1
(328, 516)	1
(329, 516)	1
(331, 516)	1
(332, 516)	2
(333, 516)	2
(334, 516)	2
(338, 514)	1
(342, 516)	1
(344, 516)	1
(356, 502)	1
(630, 1030)	1
(873, 516)	2861
(1745, 1032)	14

<기존 서울대병원>

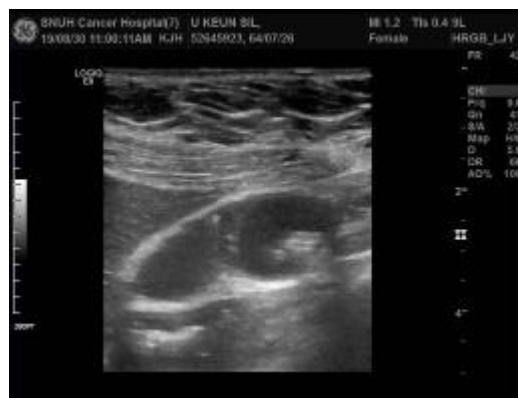
(높이, 너비)	이미지 수
(618, 1280)	3
(720, 1280)	542
(967, 1000)	3
(967, 2000)	6
(2069, 2200)	9
(2127, 2200)	9
(2127, 4400)	294

<신규 서울대병원>

담낭 용종 보고 - 서울대병원

- 기존 서울대 데이터와 신규 서울대 데이터 샘플

기존 서울대 데이터 - (873, 516)



신규 서울대 데이터 - (720, 1280)



제거 전처리 진행

담낭 용종 보고 – 실험 세팅

➤ 데이터 구성

[기존 서울대 데이터]

- 177명의 환자 데이터
- Neoplastic : 100명
- Non-neoplastic : 77명

➤ Train Set

[기존 서울대 데이터]

- 2,888장의 데이터셋
- Train Set : 177명(2,888장)

[신규 서울대 데이터]

- 86명의 환자 데이터
- Neoplastic : 15명
- Non-neoplastic : 71명

➤ Test Set

[신규 서울대 데이터]

- 866장의 데이터셋
- Test Set : 86명(868장)

➤ 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

➤ 반복 실험 30회

Top-3 Vote Method Classification Performance Table

Table . Metrics with 95% confidence interval of five Images.

Model	Metrics	AUC Mean	95% CI	
			Lower	Upper
기존 서울대 병원 데이터	AUC	0.905	0.889	0.922
	Accuracy	0.832	0.807	0.858
	Sensitivity	0.822	0.785	0.858
	Specificity	0.846	0.782	0.909
신규 서울대 병원 데이터 전처리 전 Proposed Model	AUC	0.621	0.288	0.731
	Accuracy	0.510	0.505	0.737
	Sensitivity	0.764	0.383	1.146
	Specificity	0.450	0.107	0.793
신규 서울대 병원 데이터 전처리 후 Proposed Model	AUC	0.694	0.595	0.793
	Accuracy	0.705	0.529	0.880
	Sensitivity	0.544	0.141	0.948
	Specificity	0.742	0.461	1.023

AUC, area under the curve; CI, confidence interval;