

Hello Potato World



Heo_sua_b

[테이토 논문 리뷰] GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks

ua_b 2021.3.27.05:47

[act Detection paper review]

◥ 읽었던 실험 논문 중에서 저자의 체계성과 진행 과정이 가장 눈에 잘 들어와서 정리해서 발표해보았던 논문. 의료데이터(Generative Adversarial Networks)를 사용하여 data augmentation 실험을 진행하였다.

N Augmentation

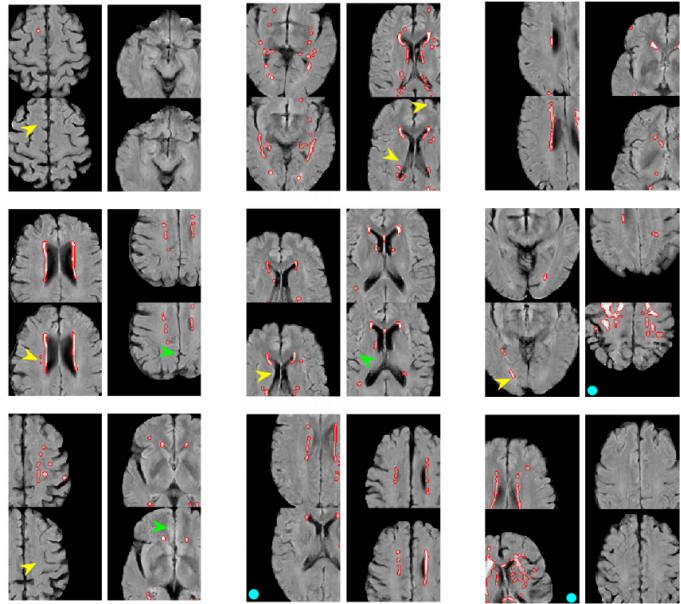
몇년간 data augmentation에 GAN을 적용시키려는 시도와 연구가 꽤 이루어지고 있다. GAN에도 이미 워낙 많은 모델들이기 때문에 다른 논문들에서는 또 여러 모델들이 사용되었지만, 여기서는 Progressive Growing of GANs(PGGAN)을 소개하였다.

논문에서 언급한 Data Augmentation의 필요성에 대해서부터 보자면, 이미지 데이터들을 다룰 때 해당 이미지 내에서 꼭 필요하지 않은 특성이 있을 것이다. 예를 들어 의료 이미지 데이터를 다룰 때는, 중요한 기관이나 조직의 위치, 모양, 크기 정보가 될 테지만 전반적인 밝기 차이나 촬영 각도 등 추출해내고자 하는 정보와 관련이 없어보인다. 이 때 쓸모있는 특성(pertinent variance), 쓸모없는 특성을 non-pertinent variance라고 하는데, 이 쓸모없는 정보를 너무 많이 유지하면 pertinance를 진단하는데 방해가 되기도 하고, overfitting의 가능성이 높아지는 등의 문제가 발생할 수 있기 때문에 사용하고 기댓값 내에서 non-pertinent variance를 제거할 수 있다. 첫 번째 방법이 데이터의 분포를 단순화시키는 것이고(intensification, cropping, registration to a standard space등의 방법이 있다), 두번째 방법이 Data augmentation 방식이고 저자는 GAN이 hand craft(직접 조정해줘야하는) feature들에 대한 필요성을 줄여준다는 장점을 살려서 augmentation에 사용하고자 했다.

Hello Potato World



Heo_sua_b



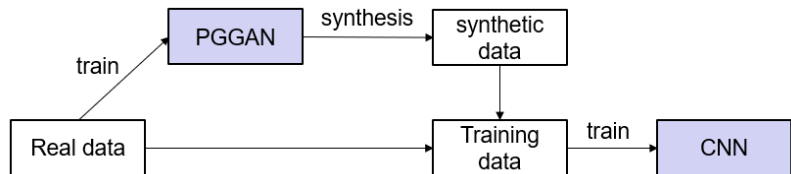
[Figure 1: 각 이미지 쌍에서 위쪽이 GAN으로 변형된 이미지, 아래가 원본 이미지]

에서 PGGAN을 사용해서 데이터를 증강하고 CNN(segmentation) 모델을 학습시킨 방법에 대해서 먼저 살펴보면, 사용하려는 데이터셋에서 총 80k개의 이미지를 뽑아내서 PGGAN을 학습시킨다.

학습된 PGGAN을 사용해서 80k개의 이미지에 대한 synthetic data를 만든다. (같은 이미지가 생성되면 안되기 때문에 PGGAN에 Gaussian Noise를 추가하는 변형을 넣었다고 한다.)

생성된 synthetic data로부터 일부를 무작위로 추출하여 기존 데이터셋에 합친다.

최종으로 생성된 Training data로 CNN(Segmentation Network)을 학습시킨다.



[Figure 2: Augmentation & Training 전체 과정]

을 평가할 때는 결과의 유사도를 측정하는 통계적 방법인 Dice Similarity Coefficient(DSC)라는 평가지표가 사용되었고, 실험과정에서 조정될 수 있는 변수는 다음과 같이 5개가 있다.

Amount of available real data: Training data는 Real data의 일부+synthetic data의 일부로 이루어 지는데, 이 때 사용되는 real data의 양

Amount of additional synthetic data: Training data에 사용되는 synthetic data의 양

Dataset: real data로 사용되는 dataset

Segmentation network: CNN에 사용되는 모델

Augmentation: Augmentation에 사용되는 방식(기존에 자주 사용되는 방식으로는 cropping, rotation, noising등이 있고 여기서는 GAN을 사용한다.)

에서는 이 5개의 변수를 조정해가며 실험결과를 비교하여 각각의 변수가 성능에 어떤 영향을 주는지 알아보려고 했다. 저자의 질문을 차례로 살펴보면 어떻게 변수를 조정했는지, 어떤 결과를 보였는지 정리해보자.



Segmentation Network에 사용되는 모델이 성능 향상에 영향을 주었는가?

CT dataset에 적합한 UNet과 Residual UNet(UResNet), 그리고 medical segmentation에서 일반적으로 많이 사용되는 DeepMedic까지 총 3개의 Segmentation Network를 사용하여 실험하였다.

Figure 3를 보면 DeepMedic을 제외한 나머지 두 개의 실험결과가 나와있다. UNet과 UResNet 사이에서 augmentation 변화 정도나 변화 방향이 비슷하기 때문에 다른 segmentation network를 사용한다고 해서 GAN augmentation의 영향이 달라지지는 않는다고 볼 수 있다.

Hello Potato World



Heo_sua_b

UNet은 어떤 segmentation network를 사용하더라도 성능 향상에 도움이 된다.

		Available data					
Additional Data		UNet			UResNet		
		100%	50%	10%	100%	50%	10%
	0%	88.9 (0.51)	86.0 (0.50)	76.9 (0.58)	86.8 (0.82)	82.7 (1.55)	72.5 (1.98)
	50%	89.2 (0.30)	87.3 (0.46)	78.6 (1.04)	86.3 (1.44)	84.3 (1.31)	74.3 (1.63)
	100%	89.3 (0.39)	86.9 (0.36)	78.4 (0.99)	86.3 (1.24)	84.1 (1.32)	74.7 (1.18)

[Figure 3: Segmentation Network]

Augmentation 방식의 차이가 성능 향상에 영향을 주는가?

더도 언급했듯이 기존에는 Data augmentation을 할 때 이미지에 기하학적 또는 밝기, 색상 등의 변환을 주는 방식을 많이 사용했는데, GAN을 사용하는 것이 이 전통적인 방법들과 비교해서 더 좋은 영향을 주는지 알아보고자 했다. 논문에서는 전통적인 augmentation의 대표 예시로 Rotation(회전) augmentation을 사용하였다. 총 4가지의 실험결과를 비교분석하였는데, 첫 번째로 augmentation을 아무것도 적용하지 않은 경우, 두 번째로 GAN augmentation을 적용했을 경우, 세 번째로 Rotation augmentation을 적용했을 경우, 마지막 네 번째로 GAN과 Rotation augmentation을 함께 적용했을 경우로 나누어서 실험

Figure 4의 결과를 보면, 어떤 방식이더라도 일단 augmentation을 적용하면 아무것도 건드리지 않았을 경우에 비해 성능이 향상되었고, 두 방식을 함께 사용했을 경우 각각 따로 적용했을 때보다 성능향상이 더 이루어진 것을 볼 수 있다. 이것은 traditional augmentation과 GAN이 각각 독립적이라는 것, 즉 각 method가 서로 다른 방식으로 동작하고, 같이 사용했을 때 시너지 효과를 일으킬 수 있다는 결론을 얻을 수 있다.

방식의 독립성으로 인해 시너지 효과를 낼 수 있다.

	Available data		
	100%	50%	10%
No augmentation	88.1 (0.32)	85.0 (0.58)	75.1 (0.60)
GAN augmentation	88.4 (0.41)	85.6 (1.33)	76.3 (1.77)
Rotation augmentation	88.9 (0.51)	86.0 (0.50)	76.9 (0.58)
GAN + Rotation augmentation	89.3 (0.39)	86.9 (0.36)	78.4 (0.99)

[Figure 4: Augmentation]

Amount of available real data가 성능 향상에 영향을 주는가?

Data Augmentation이란 방식은 학습 데이터의 양이 부족할 때 효과가 극대화될 수 있기 때문에, 사용하는 Real data의 양이 이 상황을 만들어서 비교해보고자 했다. 총 80k개의 Real data중에서 10%~90%를 무작위로 추출하여 사용했다.(라벨에서는 100%까지 나와있어서 여러번 다시 읽어보았다...아직도 의문)

Figure 5-(1)의 column에 나온 부분이 Real data의 사용량을 보여주는데, 예시로 UNet의 경우만 봐도 10%일때(첫번째 row) 100%일때는 88.9%로 매우 큰 차이를 보이고 있다. 이걸 통해서 성능향상에 가장 큰 영향을 주는 것은 Real data의 사용량이라고 확인할 수 있다.

다른 데이터셋에서 실험한 Figure 5-(2)를 보면, 100% 모두 사용했을 경우 GAN augmentation의 적용비율을 키울수록 성능이 오히려 나빠지는 것을 확인할 수 있다. 데이터의 양이 부족하지 않은 경우에는 인위적인 augmentation을 추가하는 것이 오히려 해가 될 수 있다고 확인할 수 있다.

슬 데이터의 양이 성능 향상을 결정하는 가장 중요한 요소이다.

Hello Potato World



Heo_sua_b

		Available data					
		UNet			UResNet		
		100%	50%	10%	100%	50%	10%
Additional Data	0%	88.9 (0.51)	86.0 (0.50)	76.9 (0.58)	86.8 (0.82)	82.7 (1.55)	72.5 (1.98)
	50%	89.2 (0.30)	87.3 (0.46)	78.6 (1.04)	86.3 (1.44)	84.3 (1.31)	74.3 (1.63)
	100%	89.3 (0.39)	86.9 (0.36)	78.4 (0.99)	86.3 (1.24)	84.1 (1.32)	74.7 (1.18)

		Available data		
		100%	50%	10%
Additional Data	0%	66.0 (1.26)	61.4 (2.67)	52.2 (6.65)
	50%	65.5 (1.21)	63.7 (0.69)	57.2 (4.09)
	100%	64.8 (1.34)	62.8 (1.17)	55.7 (4.26)

[Figure 5-(1), 5-(2): Amount of available real data]

*Amount of Synthetic data*가 성능 향상에 영향을 주는가?

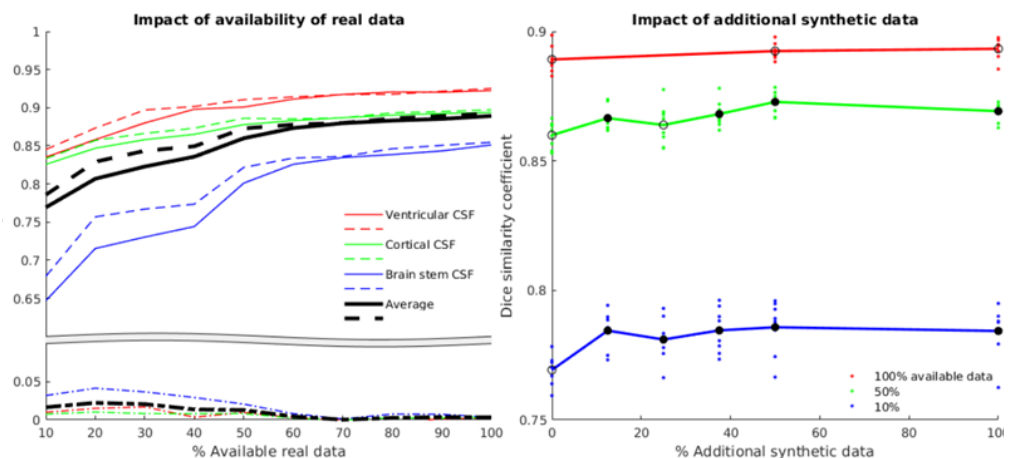
hetic data의 양이 Segmentation Network에 주는 영향을 조사하기 위해서, Training data로 합칠 Synthetic data의 양을 여러번 실험하였다. 0%~100% 사이로 조절하였는데 이 때 퍼센트 비율은 Synthetic data에 대한 비율이 아니라 real data의 퍼센트 비율로 표현되었다. 예를 들어 50%의 추가적인 data를 사용하였다고 하면, 최초 real data는 총 80k개이 10k를 더하여 총 120k의 Training data를 사용하게 되는 것이다.

결과는 위와 같은 자료인 Figure5에서 확인할 수 있으며, Synthesis과정에서 주어진 특성들과 다른 추가적인 특성들도 생길 변수가 생기기 때문에 50%내외에서 가장 큰 효과를 보이고 있다.

*Dataset*의 종류가 성능 향상에 영향을 주는가?

≡ 서로 다른 Dataset에 대해 실험해보기 위해서 CT image dataset과 FLAIR image dataset 총 두가지를 사용하였다. 주목할 수 있는 CT image인데, 이 데이터셋은 label이 cortical CSF, brain stem CSF, ventricular CSF으로 총 3개의 class로 구성되어 있다. 클래스 이미지는 순서대로 4.35:1:1.35의 비율로 존재하고 있는데, Figure 6을 참고해서 보면 알 수 있듯이 3개의 클래스가 존재하는 Brain stem CSF가 가장 큰 성능향상을 보이고 있다. (파란색 그래프) 따라서 모든 클래스가 존재하는 balanced dataset보다는 클래스 사이의 불균형이 심한 imbalanced dataset에 대해서 더 눈에 띄는 효과를 보이고 한다.

balanced dataset에 대해 더 큰 성능향상을 보인다.



[Figure 6: Dataset]



Hello Potato World



Heo_sua_b

Conclusion

한 condition 하에서 실험을 해보았을 때, 사용된 평가지표 (DSC)에 대해서 1~5%의 성능향상을 보였고, 10%의 데이터만 데이터 부족) 상황에서 가장 큰 향상을 보였다.

이 굉장히 체계적으로 짜여져서 이루어졌고, 결과도 전부 보기 좋게 정리되어 있어서 읽기가 편했던 것 같다. 제일 흥미로웠던 기존의 traditional augmentation 방식들과 GAN이 아예 다른 방향으로 동작된다는 점이었고, 조금 더 깊게 수식적으로 새로운 아이디어를 얻을 수 있지 않을까 라는 생각이 든다.

References

Christopher Bowles et al, GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks, 2018

이 글은 학생 혼자 읽고 기록하려고 남기는 리뷰입니다 수정할 부분은 알려주세요 🍌

7

구독하기

더 리뷰 🍌 > Data Augmentation' 카테고리의 다른 글

이토 논문 리뷰] GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks (0)	2021.0
이토 논문 리뷰] Cascade Eff-B7 NAS-FPN : Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation (0)	2021.0
이토 논문 리뷰] Learning Data Augmentation Strategies for Object Detection (4)	2020.1

더 리뷰 🍌 / Data Augmentation' Related Articles

[포테이토 논문 리뷰]
 Cascade Eff-B7 NAS-FPN : Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation
 2021.03.17



Hello Potato W... 구독하기

omments



여러분의 소중한 댓글을 입력해주세요

Secret Send

Prev 1 ... 12 13 14 15 16 17 18 19 20 Next

Hello Potato World



Heo_sua_b

Blog is powered by [kakao](#) / Designed by [Tistory](#)