



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

광각 안저사진을 이용한  
적대적 생성 신경망 합성 결과 비교 연구

제주대학교 대학원

융합교육소프트웨어학과

고 아 라

2020년 8월

# 광각 안저사진을 이용한 적대적 생성 신경망 합성 결과 비교 연구

지도교수 조 정 원

고 아 라

이 논문을 이학 석사학위 논문으로 제출함

2020년 6 월

고아라의 이학 석사학위 논문을 인준함

심사위원장 \_\_\_\_\_ 박 찬 정



위 원 \_\_\_\_\_ 김 성 백



위 원 \_\_\_\_\_ 조 정 원



제주대학교 대학원

2020 년 8 월



<국문초록>

## 광각 안저사진을 이용한 적대적 생성 신경망 합성 결과 비교 연구

고 아 라

제주대학교 대학원 융합교육소프트웨어학전공

지도교수 조 정 원

최근 의료현장에서는 의료 촬영술의 발전으로 기존의 일반 안저촬영기에서 촬영이 더욱 간편해지고 촬영 범위가 넓어진 광각 안저촬영기로 교체되고 있다. 그러나 데이터 수집 과정의 어려움으로 인해 현재 대부분 안과 질환 관련 딥러닝 연구는 여전히 기존 일반 안저사진을 기반으로 이루어지고 있다.

딥러닝의 성능은 데이터 수량과 연관이 있어 데이터가 부족한 경우 데이터 증강을 사용해 데이터를 증폭하여 성능을 향상한다. 데이터 증강에는 이미지에 회전, 반전 등을 적용한 일반적인 데이터 증강 방법이 있으며, GAN을 사용하여 새로운 이미지를 합성하는 방법이 존재한다.

GAN은 다양한 모델이 존재하며 각 모델의 성능을 평가하기 어렵다. 모델에 사용된 데이터 세트가 다르며 적절한 평가 척도가 존재하지 않기 때문이다. 또한 GAN 모델 연구에 사용된 데이터가 실제 합성에 사용되는 데이터의 해상도, 복잡도를 반영하지 못한다. 따라서 본 논문에서는 다양한 GAN 모델을 이용해 광각 안저사진을 합성하고 결과를 비교·분석한다.

광각 안저사진은 제주대학병원 안과와 협업을 통해 당뇨망막병증으로 한정 지어 데이터를 수집 및 라벨링하였다. 본 연구에 사용된 GAN 모델은 의료 영상 합성 연구에서 자주 사용되는 DCGAN, WGAN-GP, BEGAN을 이용하였다. 각 모델의

하이퍼 파라미터는 일부는 동일하게 설정하고, 나머지 값들은 모델이 발표된 논문을 기준으로 설정하였다.

합성된 안저사진은 정성적, 정량적 평가를 진행하였으며, 정성적 평가는 합성 결과에서 시신경유두, 황반 등의 관찰 가능 여부를 판단하였다. 정량적 평가는 FID와 이미지 품질 평가 척도를 이용하였다. FID의 경우 BEGAN 모델을 이용한 경우 가장 좋은 평가를 얻었으며, 이미지 품질 평가의 경우 WGAN-GP 모델이 결과가 좋았다. 정성적 평가와 정량적 평가 후 결과를 비교하였을 때 BEGAN이 가장 적절하다 판단하였다. 또한, BEGAN의 경우 훈련을 계속 반복하였을 때 실제 안저사진과 비슷한 결과를 얻을 수 있었다. 데이터 세트의 특성에 따라 BEGAN의 하이퍼 파라미터를 조정한다면 훨씬 더 정교한 결과를 얻을 수 있을 것으로 기대된다.

## <제목 차례>

I. 서론 .....	1
1. 연구 배경 및 목적 .....	1
2. 연구 내용 및 방법 .....	2
II. 이론적 배경 .....	4
1. 안저검사와 당뇨망막병증 .....	4
2. 광각 안저사진 활용 연구 .....	6
3. 안저사진 합성 연구 .....	8
III. 관련 연구 .....	12
1. GAN .....	12
1.1. DCGAN .....	14
1.2. WGAN과 WGAN-GP .....	16
1.3. BEGAN .....	17
2. 평가 척도 .....	19
IV. 합성 모델 설계 및 결과 비교 .....	22
1. 실험 데이터 .....	22
2. GAN 모델 및 실험 환경 .....	24
3. 학습 단계별 데이터 분포 변화 .....	28
4. 합성 결과 비교 .....	36
4.1. 정성적 평가 .....	36
4.2. 정량적 평가 .....	39

V. 결론 .....	45
[참고문헌].....	50
[Abstract].....	54
[부록].....	56

## <표 차례>

<표 1> 광각 안저사진 연구 요약 .....	7
<표 2> GAN 모델에 사용된 데이터 세트 .....	14
<표 3> 개발 환경 .....	24
<표 4> 각 모델별 안저사진 요소 평가 .....	38
<표 5> 모델별 FID 점수 .....	40
<표 6> 실제 데이터 기반 FID 백분율 .....	40
<표 7> 모델별 MSE 점수 .....	41
<표 8> 모델별 MAE 점수 .....	42
<표 9> 모델별 PSNR 점수 .....	43
<표 10> 모델별 SSIM 점수 .....	43
<표 11> 모델별 정량적 평가 결과 (정상 데이터) .....	48
<표 12> 모델별 정량적 평가 결과 (병증 데이터) .....	48



## <그림 차례>

[그림 1] 안저촬영에 따른 안저사진 비교 .....	5
[그림 2] Vanilla GAN을 이용한 이미지 합성 과정 .....	8
[그림 3] image-to-image translation을 이용한 이미지 합성 과정 .....	9
[그림 4] 혈관 사진과 안저사진 .....	9
[그림 5] Biswas의 연구 결과 예시 .....	11
[그림 6] DCGAN 구조 .....	15
[그림 7] WGAN - 실제 이미지와 생성된 이미지의 확률분포 .....	16
[그림 8] WGAN - 값에 따른 확률분포 거리 .....	17
[그림 9] BEGAN 구조 .....	19
[그림 10] 광각 안저사진에서 검출 가능한 증상 .....	22
[그림 11] 당뇨망막병증 중증도에 따른 안저사진 .....	23
[그림 12] 구현한 DCGAN 구조 .....	25
[그림 13] 구현한 WGAN-GP 구조 .....	26
[그림 14] 구현한 BEGAN 구조 .....	27
[그림 15] 정상 데이터와 병증 데이터의 분포 .....	28
[그림 16] DCGAN 학습 과정 - 정상 데이터 사용 .....	29
[그림 17] DCGAN 학습 과정 - 병증 데이터 사용 .....	30
[그림 18] WGAN-GP 학습 과정 - 정상 데이터 사용 .....	31
[그림 19] WGAN-GP 학습 과정 - 병증 데이터 사용 .....	31
[그림 20] BEGAN(0.3) 학습 과정 - 정상 데이터 사용 .....	32
[그림 21] BEGAN(0.3) 학습 과정 - 병증 데이터 사용 .....	33
[그림 22] BEGAN(0.5) 학습 과정 - 정상 데이터 사용 .....	33
[그림 23] BEGAN(0.5) 학습 과정 - 병증 데이터 사용 .....	34
[그림 24] BEGAN(0.7) 학습 과정 - 정상 데이터 사용 .....	34
[그림 25] BEGAN(0.7) 학습 과정 - 병증 데이터 사용 .....	35
[그림 26] DCGAN 합성 결과 (20,000 epoch) .....	36

[그림 27] WGAN-GP 합성 결과(20,000 epoch) .....	37
[그림 28] BEGAN 합성 결과 (20,000 epoch) .....	37
[그림 29] BEGAN 합성 결과 (증상이 있는 데이터 사용) .....	39
[그림 30] 정상 데이터를 이용한 합성 결과 데이터 분포 .....	46
[그림 31] 병증 데이터를 이용한 합성 결과 데이터 분포 .....	46

## <수식 차례>

(1) GAN 목적함수 .....	13
(2) Wasserstain Distance .....	16
(3) Diversity ratio .....	18
(4) BEGAN 목적함수 .....	18
(5) IS .....	19
(6) FID .....	19
(7) MSE .....	20
(8) MAE .....	20
(9) PSNR .....	20
(10) SSIM .....	21

# I. 서론

## 1. 연구 배경 및 목적

눈의 건강을 확인하는 데 있어 가장 기본적인 검사는 안저검사이다. 안저검사를 통해 망막의 중심부와 주변부를 관찰하여 다양한 질병 검출이 가능하다. 안저검사를 위한 안저촬영법은 관찰 가능한 범위에 따라 일반 촬영과 광각 촬영으로 나뉜다. 일반 촬영의 경우 30-50°를 관찰할 수 있으나, 광각 촬영은 200° 범위의 망막 주변부를 더 넓게 관찰할 수 있어 병증의 진단율을 높이는 방법으로 주목받고 있다[1]. 광각 안저사진을 이용하여 다양한 질병을 관찰할 수 있어 각 병원에서의 안저검사가 광각 안저검사로 바뀌고 있다.

최근 다양한 산업에 인공지능 기술이 접목되고 있으며 그중에서도 의료 부분이 주목받고 있다. 그중에서도 다양한 의료 영상을 이용하여 딥러닝 연구들이 진행되고 있으며, 안저검사의 경우 대표적으로 당뇨망막병증(당뇨망막증, Diabetic Retinopathy)이 있다.

당뇨망막병증의 경우 발병되고 난 뒤 완전한 치료는 불가능하므로 위험을 낮추기 위해 정기적인 검진을 통한 조기 진단이 필수이다[2]. 정확하고 빠른 진단을 위해 딥러닝을 이용한 병증 판별 연구들이 진행되고 있으나 연구를 위해 다량의 데이터가 필요하다. 그러나 의료 영상의 데이터는 수집이 어렵고 연구 목적에 맞춰 데이터를 라벨링하는 과정에서 상당한 시간과 비용이 소모된다. 또한, 윤리적 이슈로 인해 데이터의 공유는 거의 이뤄지지 못하고 있다. 대부분 연구가 데이터 수집 등의 어려움으로 인하여 공개된 데이터 세트인 일반 안저사진을 이용하고 있으며, 이러한 시스템은 광각 안저사진을 이용하여 병증을 진단할 수 없다는 문제점이 있다.

딥러닝의 성능은 데이터의 양과 밀접한 관련이 있다. 수집된 데이터의 수가 적은 상태에서 진행되는 연구들은 데이터 증강 방법을 사용하여 데이터의 양을 늘려 연구를 진행한다. 데이터를 증강하는 방법은 이미지의 경우 회전, 반전 등의 효과를

적용한 일반적 방법이 존재하며, 생성 모델 중 하나인 GAN을 이용해 새로운 데이터를 생성하는 방법이 존재한다. 일반적 증강 방법의 경우 적용하기 쉬우나, 병증 유무에 따른 데이터 편향 문제를 해결하지 못한다. 이에 반해 GAN을 사용하는 경우 방법을 찾기 어려우나 적절한 모델을 구현한다면 새로운 데이터를 생성할 수 있어 편향 문제를 해결할 수 있다.

GAN을 이용해 안저사진을 합성하는 연구들은 모두 일반 안저사진을 이용해서 진행하였다. 또한, 선명한 이미지 합성을 위해 대부분 연구에서 안저사진 외에 추가로 혈관 사진(Vessel Tree)을 이용하였다. 그러나 일부 연구에서 안저사진만을 이용해 연구를 진행하였고, 실제 안저사진과 비슷한 이미지를 생성하는 데 성공하였다. 해당 연구에서 안저사진 합성의 가능성을 발견하였으며 본 논문에서는 다양한 GAN 모델을 이용하여 광각 안저사진을 합성하고, 결과를 비교·분석하고자 한다.

## 2. 연구 내용 및 방법

본 연구를 위해 제주대학병원 안과와 협업하여 광각 안저사진을 수집하였다. 광각 안저사진의 경우 관찰 범위가 넓어 다양한 증상 검출이 가능하지만, 당뇨망막병증으로 한정 지어 데이터를 수집·라벨링 하였다.

GAN이 발표된 이후 다양한 파생 모델이 등장하였고, 각 모델에서 합성된 데이터가 다르다. 또한 합성된 결과를 판단할 수 있는 지표가 존재하지 않아 모델 간 성능을 비교·분석하는데 어려움이 존재해 어느 모델이 더 성능이 좋은지 판별하기 힘들다. 따라서 의료 영상 합성에서 사용되었던 모델을 선정하여 각 모델의 합성 결과를 이용해 모델의 성능을 비교하고, 광각 안저사진 합성에 적절한 모델을 제시하고자 한다.

본 연구에서 비교 대상으로 선정된 모델은 DCGAN, BEGAN, WGAN-GP이다. DCGAN과 WGAN의 경우 의료 영상 합성 연구에서 자주 사용되는 모델이며, WGAN보다 WGAN-GP가 학습 안정성이 높아 WGAN-GP를 선택하였다. GAN은 동일한 환경을 구성하여 비교하는 연구가 어려워 각 모델의 구현은 제안된 논문을

기준으로 진행하였다. 또한, 세 개의 모델을 공통으로 세팅한 경우 외에도 추가로 진행한 실험의 하이퍼 파라미터와 결과를 부록으로 제시하였다.

합성된 이미지를 평가할 수 있는 적절한 평가 척도가 존재하지 않아 학습 과정 동안의 커널 밀도 추정을 통해 데이터 분포를 살펴보았다. 또한, 합성된 결과를 이용해 정성적 비교와 정량적 비교를 나눠 진행하였다. 정성적 비교에서는 합성 결과를 제시하였고, 안저사진에서 관찰되어야 하는 전체적인 윤곽, 시신경유두, 황반, 혈관, 증상의 유무를 비교하였다. 정량적 비교의 경우 가장 많이 사용되는 평가 척도인 FID와 이미지 품질 평가 척도들(MSE, MAE, PNSR, SSIM)을 이용하였다. 비교군이 존재하지 않아 실제 데이터와 실제 데이터의 점수들을 측정하였으며, 해당 점수를 기준으로 평가를 진행하였다.

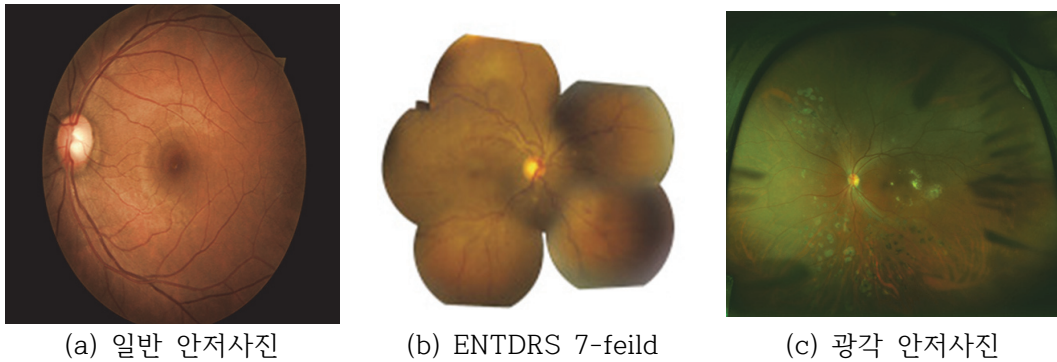
## II. 이론적 배경

이 장에서는 본 논문의 이론적 배경을 기술한다. 첫 번째로 본 연구에 사용된 데이터인 안저사진에 관해 살펴본다. 두 번째로 광각 안저사진을 활용한 딥러닝 연구들을 살펴보고, 마지막으로 합성 연구 진행을 위해 일반 안저사진 합성 연구에 대해 살펴본다.

### 1. 안저검사와 당뇨망막병증

망막을 관찰하기 위해서는 안저촬영이 필요하며, 안저촬영은 관찰 가능한 망막의 범위에 따라 일반 안저촬영과 광각 안저촬영으로 나뉜다. 일반 안저촬영의 경우 30-50° 범위를 관찰할 수 있어 약 15% 정도의 망막을 볼 수 있다. 광각 안저촬영의 경우 200° 범위를 관찰할 수 있으며, 이는 약 80% 정도의 망막에 해당한다. 관찰 가능한 망막의 범위가 넓어 일반 안저촬영 보다 광각 안저촬영을 진행하는 경우 병증 진단에 유리하다. 이 외에도 일반 안저촬영의 경우 동공을 확대하기 위해 산동제를 투여해야 하는데, 이 경우 동공이 확대되어 있으므로 검사 후 바로 일상 생활을 하기에 무리가 있을 수 있다. 그러나 광각 안저촬영의 경우 산동제를 투여하지 않기 때문에 검사 후 바로 생활 할 수 있다.

광각 안저촬영법이 등장하기 이전에는 좀 더 넓은 망막을 관찰하기 위해 일반 안저사진 7장을 연결하여 합성하는 방법(ETDRS 7-field)을 이용하였으나 합성 시간이 오래 걸리고 다양한 범위를 촬영해야 하는 단점이 존재한다. ETDRS 7-field와 광각 안저사진을 비교하면 광각 안저사진을 이용해 등급을 진단할 때 좀 더 심각한 등급이 진단되기도 하였다[3]. 일반 안저사진과 ETDRS 7-field, 광각 안저사진의 예는 [그림 1]과 같다.



[그림 1] 안저촬영에 따른 안저사진 비교

광각 안저 촬영은 주변부 망막을 간편하게 확인할 수 있어, 기존의 검사에서 발견하지 못한 병변들을 추가로 찾아내 치료할 수 있는 장점이 있다[4]. 또한, 증식성 당뇨망막병증의 경우 신생 혈관을 관찰하는 것이 중요하며 ENTDRS 7-field와 광각 안저사진을 비교하였을 때 광각 안저사진을 이용한 신생 혈관 추적이 좀 쉽다.

당뇨망막병증은 당뇨병의 합병증 중 하나로 당뇨병으로 인해 말초신경에 순환장애가 생기는 병증이다. 당뇨망막병증으로 인해 망막에 있는 혈관이 쉽게 터질 수 있으며, 혈관이 터지게 되면 당이 포함된 혈액이 흘러 증상이 심해지는 경우 황반이 손상되어 실명으로 이어질 수 있는 질병이다. 또한, 성인 실명 원인 1위를 차지할 만큼 위험한 질병이다[5].

당뇨망막병증은 병증이 심각해지기 이전에는 증상이 잘 나타나지 않아 정기적인 검진이 필수이다. 발병 후 완전한 치료가 불가능하며 혈당 조절을 통해 발병을 늦추거나 발병 후 병증의 진행 속도를 늦추는 방법만 존재한다. 따라서 조기 진단이 중요한 질병 중 하나이다.

당뇨망막병증은 혈관 생성에 따라 증식성(Proliferative)과 비증식성(Non-Proliferative)으로 나눌 수 있고, 비증식성은 중증도에 따라 Mild, Moderate, Severe로 나눌 수 있다. 각 중증도는 관찰할 수 있는 미세혈관류, 망막내출혈 등을 통해 판단한다. 이러한 증상은 혈관이 존재하는 다양한 범위에서 관찰 가능하며, 관찰 범위가 넓을수록 정밀한 판단이 가능하다.



## 2. 광각 안저사진 활용 연구

광각 안저사진을 활용한 연구는 최근 들어 발표되기 시작하였다. 각 연구를 살펴보면 만장이 넘는 데이터를 사용한 연구도 있지만 몇백 장의 사진을 이용해 진행된 연구도 있다. 또한, 모든 데이터 세트가 병원과 협력을 통해 얻은 데이터이기 때문에 공개하지 않았다. 데이터의 라벨링은 병원의 망막 전문가(retinal specialist) 또는 안과 전문의(ophthalmologist)를 통해 진행되었다.

다양한 안과학적 병증을 진단하는데 광각 안저사진을 활용한 연구들이 진행되었으며, 이러한 연구들이 부족한 광각 안저사진 데이터를 극복하는 방법으로 사용한 데이터 증강방법과 연구에서 사용된 이미지 사이즈 측면에서 관련 연구들을 분석해 보았다.

Toshihiko et al(2019)의 연구[6]에 따르면 광각 안저사진을 이용해 치료가 미숙한 증식성 당뇨망막병증을 검출하였는데, 378장의 이미지를 사용하였다. 256\*192 사이즈로 조절하여 사용했으며 0-1 사이의 값으로 정규화하였다. 데이터 증강 방법은 밝기 조정, 감마 보정, 히스토그램 평활화(Histogram equalization), 노이즈 추가, 반전을 이용하였고 훈련 데이터를 18배 증강시켰다.

해당 연구의 한계점은 실제 의료 환경을 반영하지 못했다는 것이다. 송경두(2019)에 의하면 의료 영상을 이용한 연구는 실제 환경을 반영한 데이터 확보가 중요하다고 지적하였다[7]. 해당 연구에서는 치료받지 않은 증식성 당뇨망막병증으로 한정하였다. 증식성 당뇨망막병증의 경우 증상이 심각하기 때문에 치료받지 않은 환자를 찾기 어려우며, 정상 데이터와 비교하였을 때 병증의 특징이 뚜렷하기 때문에 다른 중증도에 비해 검출이 쉽다.

당뇨망막병증 외에도 특발성 황반원공(Idiopathic macular hole)[8], 류마티스 망막 박리(Rhegmatogenous retinal detachment)[9], 색소성 망막염(Retinitis pigmentosa)[10], 망막 분지 정맥폐쇄(Branch retinal vein occlusion)[11] 진단 연구들이 있다. 류마티스 망막 박리 검출의 경우 96\*96 사이즈로 이미지를 줄여 사용하였으며, 그 외의 연구에서는 256\*192 사이즈의 이미지를 사용하였다. 각각의 연구에서는 910, 831, 373, 466장을 데이터로 사용하였다. 특발성 황반원공 검출 연구의 경우 약 8배 훈련 데이터를 증강하였으며, 그 외의 연구에서는 18배 훈련 데이터를 증강하였다.

이 외에도 Hemorrhage screening[12]과 망막격자변성(Retinal lattice degeneration) 검출[13]이 있다. Hemorrhage screening 연구의 경우 약 1만 7천 장의 사진을 사용했으며, 512\*512 사이즈로 줄이고 표준화와 정규화(0~1)를 진행하였다. 데이터 증강을 이용해 약 5만 5천 장의 이미지를 만들었으며, 0.8 ~ 1.6 범위로 무작위의 밝기 이동과 최대 90° 회전, 상하 반전, 좌우 반전을 이용하였다. 망막격자변성의 경우 훈련 데이터를 3가지로 나눠 연구를 진행하였으며, 첫 번째는 원본의 데이터 세트, 두 번째는 밝기 이동, 회전, 반전을 이용한 증강, 마지막은 히스토그램 평활화를 진행하여 각 경우를 비교하였다.

모든 연구에서 일반적 데이터 증강 방법을 이용해 데이터의 수를 늘렸다는 것을 알 수 있었으며, 반대로 데이터의 수가 부족하므로 증강 방법을 이용했다고 추론할 수 있다. 또한, 모든 연구에서 병원과 협업을 통해 데이터를 수집하였기 때문에 데이터가 공개되어 있지 않다. 그리고 고해상도의 광각 안저사진을 일정 수준으로 줄여 진단 연구에 사용된다는 것을 알 수 있었다. 해당 연구의 요약은 <표 1>과 같다.

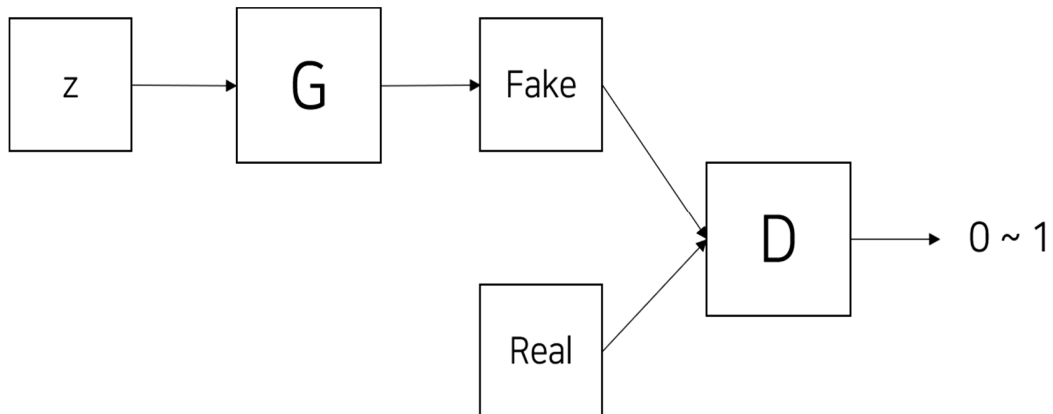
<표 1> 광각 안저사진 연구 요약

병증 종류	사용된 총 이미지 수	데이터 증강 방법	훈련 데이터 증강 비율	사용된 이미지 사이즈
증식성 당뇨망막병증[6]	378	밝기 조정, 반전, 감마 보정, 노이즈 추가, 히스토그램 평활화	18배	256 * 192
특발성 황반원공[8]	910	밝기 조정, 반전, 감마 보정, 노이즈 추가, 히스토그램 평활화	약 8배	256 * 192
류마티스 망막 박리[9]	831	밝기 조정, 반전, 감마 보정, 스무딩, 노이즈 추가	18배	96 * 96
색소성 망막염 [10]	373	밝기 조정, 반전, 감마 보정, 노이즈 추가, 히스토그램 평활화	18배	-

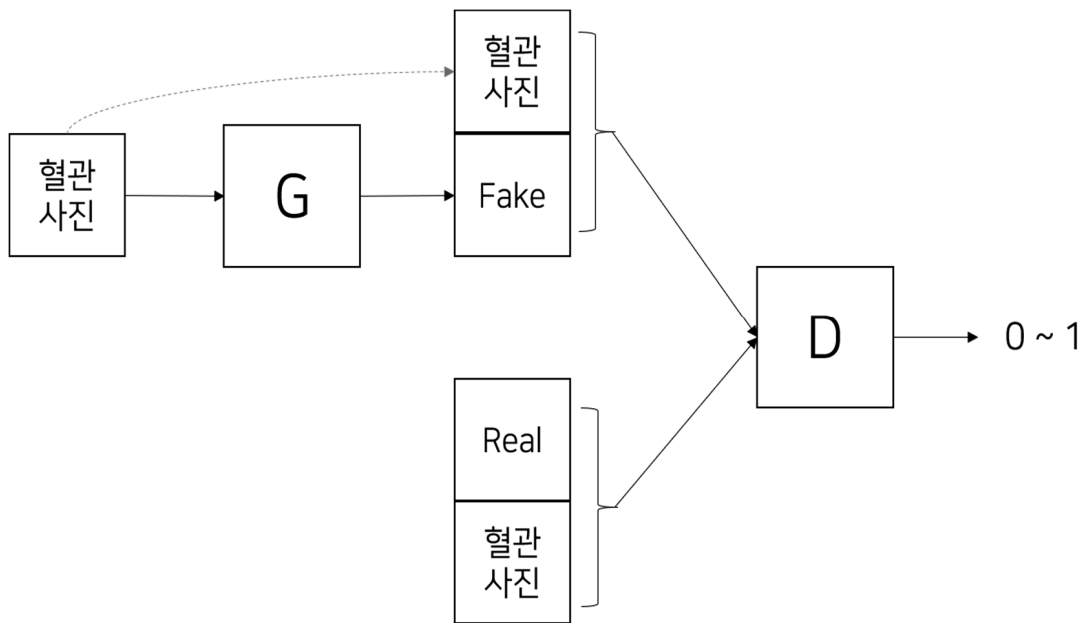
망막분지 정맥폐쇄[11]	466	밝기 조정, 반전, 감마 보정,노이즈 추가, 히스토그램 평활화	18배	256 * 192
Hemorrhage Screening[12]	16,827	좌우 및 상하 반전, 회전(최대 90°), 밝기 조정	약 5배	512 * 512
망막격자변성[13]	5,606	1) 좌우 및 상하 반전, 회전(최대 90°), 밝기 조정 2) 히스토그램 평활화	약 5배	512 * 512

### 3. 안저사진 합성 연구

GAN을 이용한 일반 안저사진 합성 연구는 대부분 혈관 사진(vessel tree)을 이용한다. 혈관 사진 데이터를 추가로 이용하고, image-to-image translation을 사용해 혈관 사진에 색, 텍스처 등을 입혀 안저사진을 완성하였다. Vanilla GAN의 경우 이미지를 합성하는 과정이 [그림 2]와 같고, image-to-image translation을 이용하여 이미지를 합성하는 과정은 [그림 3]과 같다.



[그림 2] Vanilla GAN을 이용한 이미지 합성 과정



[그림 3] image-to-image translation을 이용한 이미지 합성 과정

image-to-image translation의 경우 여름 사진을 겨울 사진으로 변환하거나 사진을 특정 화가의 화풍으로 변환, 강아지를 고양이로 변환하는 등 두 데이터 세트 간의 스타일을 변환시키는 방법을 의미한다. 일반 안저사진의 경우 혈관 사진과 안저사진이 함께 존재하는 데이터 세트가 있으며, 대부분 연구에서는 해당 데이터 세트들을 이용한다. 대표적으로 DRIVE[14]와 STARE[15] 데이터 세트가 존재한다. DRIVE 데이터 세트의 경우 안저사진과 혈관 사진이 각각 40장씩 존재하며, STARE 데이터 세트는 안저사진이 20장, 혈관 사진이 20장 존재한다. 각 데이터의 수량이 부족하여 기존의 데이터 세트를 이용해 새로운 혈관 사진을 합성하는 예도 존재한다. [그림 4]는 DRIVE 세트의 데이터 중 하나의 예시이다.



[그림 4] 혈관 사진과 안저사진 (출처 : DRIVE[14])

Pedro Costa et al(2017) 연구[16]에서는 혈관 사진과 안저사진이 쌍으로 존재하는 데이터의 수량이 적어 U-Net을 이용해 혈관 분할(segmentation)을 추가로 진행하였다. 해당 연구에서는 안저사진과 혈관 사진 모두 합성하였고, 오토인코더(auto-encoder)를 이용해 혈관 사진을 합성하며, GAN을 이용해 안저사진을 합성하였다. 합성된 이미지의 평가는 ISC(Image Structure Clustering)을 이용하였으며, 실제 U-Net에 합성된 안저사진을 넣어 분할이 잘 되었는지 ROC 곡선을 이용해 결과를 나타냈다. ISC는 이미지 분할을 정량적으로 평가할 수 있는 척도로서 실제 U-Net을 이용하여 혈관 분할을 진행하였기 때문에 해당 척도를 이용하여 평가를 진행하였다.

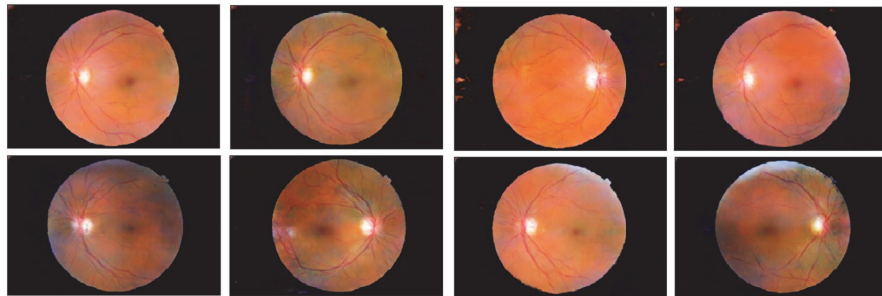
John T.Guibas et al(2017) 연구[17]에서는 geometry와 photorealism으로 나눠 생성기를 구성하였다. geometry 부분에서는 DCGAN을 이용해 혈관 사진을 합성하고, photorealism 부분에서는 image-to-image translation을 이용해 안저사진을 합성하였다. 평가는 U-Net을 이용하여 합성된 이미지에서 실제 혈관이 제대로 분할되는지 진행하였고, DRIVE 데이터와 실제 생성된 이미지 50쌍을 이용해 F1 점수를 구하였다. 추가로 실제 이미지와 합성된 이미지의 분포 차이를 구하기 위해 KL divergence score를 이용하였다.

He Zhao et al(2018)[18]와 Talha Iqbal et al(2018)[19]는 혈관 사진을 합성하지 않고 연구를 진행하였다. Zhao는 style loss와 content loss를 만들었으며, 적은 수의 데이터 샘플을 이용해 실제와 비슷한 안저사진을 만들었다. 안저사진을 합성한 후 VGG 네트워크를 이용해 style transfer를 만들었으며, style input에 따라 이미지가 변하도록 만들었다. style input이란 기존의 안저사진 데이터 세트마다 안저사진의 색과 텍스처가 다르므로 Kaggle, DRIVE, STARE를 선택하여 입력으로 줄 수 있다. 평가는 분할 모델(segmentation model)을 이용해 F1 점수, 특이도, 민감도를 측정하여 진행했으며, 합성 예제를 보였을 때, 400장 중 40장의 사진만이 실패한 경우라고 공개하였다.

Iqbal도 비슷하게 스타일 변환(style transfer)을 진행했으며, 평가는 분할 모델을 이용한 Dice coefficients(F1-score)와 판별기의 AUC ROC(Area Under Curve for Receive Operating Characteristics), AUC PR(Area Under Curve for Precision and Recall)을 이용하였다.

Sangeeta Biswas et al[20]는 혈관 사진을 이용하지 않고 안저사진만을 이용하여 연구를 진행하였다. 해당 연구에서는 DCGAN을 이용해 안저사진을 합성하고, 생성기와 판별기의 균형을 맞추기 위해 생성기와 판별기의 반복 훈련 횟수( $r$  - 생성기,  $k$  - 판별기)를 의미하는 하이퍼 파라미터를 추가하여 이용하였다. 합성된 이미지의 평가는 SSIM(Structural Similarity)을 이용하였으며, 진짜-진짜, 가짜-진짜, 가짜-가짜 그룹을 만들어 SSIM 값을 측정하였다. SSIM 점수 외에 추가적인 방법은 사용하지 않았으며, 반복 훈련 횟수에 따른 안저사진 합성 결과를 보여주고 있다. [그림 5]는 Biswas 연구 결과 중 일부로 첫 번째 행의 사진은  $k$ 의 값이 5인 경우이고 두 번째 행의 사진은  $k$ 의 값이 6인 경우이다.  $r$ 의 값은 모두 1로 설정되었다.

대부분 연구에서 혈관 사진을 이용하였지만 Biswas의 연구에서는 안저사진만을 이용하였다. 혈관 사진 없이 안저사진을 합성할 수 있다는 가능성을 발견하였다.



[그림 5] Biswas의 연구 결과 예시[20]  
(첫 번째 행 :  $k=5$  / 두 번째 행 :  $k=6$ )

### III. 관련 연구

이 장에서는 본 연구에 사용된 방법론을 살펴보기로 한다. 먼저 GAN의 주요 특징과 본 연구에서 사용된 DCGAN, WGAN-GP, BEGAN에 대해 알아본다. 이후 합성 결과를 비교하기 위한 정량적 평가 척도에 대해 살펴본다. 정량적 평가 척도는 GAN의 합성 결과를 비교할 수 있는 IS와 FID가 있으며, 이 외에도 이미지 품질 평가 척도를 이용하여 합성 결과를 판단한다. 따라서 이 장에서는 IS, FID, 이미지 품질 평가 척도를 살펴본다.

#### 1. GAN

GAN(Generative Adversarial Network)[21]은 2014년 Ian Goodfellow가 제안했으며, 생성기(Generator)와 판별기(Discriminator)로 구성된 신경망이다. 발표 이후 Yann Lecun 교수는 “10년 안에 나온 딥러닝 아이디어 중 최고의 생성 모델”이라며 찬사를 보냈으며, 이후 수많은 GAN 모델이 발표되었다.

GAN의 생성기와 판별기는 위조지폐범과 경찰관으로 비유된다. 위조지폐범은 실제 지폐와 유사한 지폐를 생성하고, 경찰관은 지폐 감별 훈련을 통해 지폐의 위조 여부를 밝혀낸다. 처음 위조지폐범이 만든 지폐는 허술하지만, 경찰관의 감별 결과를 통해 위조지폐범의 진행 방향을 결정할 수 있다. 이러한 과정을 대립이라 하며, 위조지폐범과 경찰관의 대립이 진행될수록 위조지폐는 실제 지폐와 유사해진다.

GAN의 최종 목표는 생성기와 판별기의 대립으로 인해 실제와 비슷한 데이터를 생성하는 것이다. 생성기는 주어진 입력을 실제 데이터의 분포와 유사하게 만들며, 판별기는 주어진 입력 데이터가 진짜인지 가짜인지 판별하게 된다.

실제 데이터를  $x$ 라고 하고 생성기에 주어진 입력을  $z$ 라고 했을 때, 판별기는  $D(x) = 1$ ,  $D(G(z)) = 0$ 으로 출력하는 것을 목표로 학습한다. 이에 반해 생성기는 실제 데이터의 분포를 학습하여 실제와 유사한 데이터를 생성하고,  $D(G(z)) = 1$ 이

되는 것을 목적으로 한다. 따라서  $D(G(z))$ 의 값이 증가할수록  $\log(1 - D(G(z)))$ 의 값이 감소하게 된다. 이러한 GAN의 목적 함수를 수식으로 나타내면 수식 1과 같으며, 이는 최소-최대 게임과 비슷한 양상을 보인다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

두 신경망이 대립적으로 훈련하기 위해서는 두 신경망의 균형을 적절히 유지해야 한다. 그러나 실제 데이터와 가짜 데이터를 판별하는 일보다 실제와 유사한 데이터를 생성하는 일이 더 어려워서 대부분 판별기의 성능이 먼저 더 좋아진다. 생성기와 판별기의 훈련 도중 어느 한 모델이 먼저 수렴하는 경우 한쪽의 성능 향상이 어려워진다. 또한, 생성기가 판별기를 속이기 위해 다양한 분포의 데이터가 아닌 특정 분포의 데이터만 생성하는 모드 붕괴(Mode Collapse) 문제가 발생한다. 이런 문제점으로 인해 훈련하기 어려운 모델로 손꼽히며, 안정적인 학습 진행을 위해 다양한 GAN 모델이 존재한다.

GAN 모델의 비교 연구가 필요한 이유는 각 모델 성능 평가를 위해 사용된 데이터가 다르기 때문이다. <표 2>는 일부 GAN 모델에서 사용된 데이터 세트를 나타낸다. 해당 표를 살펴보면 모델의 성능 향상을 보이기 위해 사용된 데이터 세트가 각각 다르다는 것을 알 수 있다. 또한, 현재 GAN 연구에서 사용되고 있는 모델들은 2017년 이전에 발표되었고, 현재 많이 사용되고 있는 정량적 평가척도인 FID(Frechet Inception Distance)가 2017년에 발표되었다. 따라서 정량적 평가 방법을 이용하여 각 모델을 평가하기 어렵다.



<표 2> GAN 모델에 사용된 데이터 세트

모델 종류	사용된 데이터 세트
GAN	MNIST, CIFAR-10
LSGAN	MNIST, LSUN-bedrooms(112*112)
DCGAN	LSUN-bedrooms(64*64)
WGAN	LSUN-bedrooms(64*64)
WGAN-GP	LSUN-bedrooms(128*128)
EBGAN	CIFAR-10
BEGAN	CelebA

이 외에도 GAN 훈련에 사용된 데이터가 실제 GAN을 이용해 합성하고자 하는 데이터의 특성을 반영하지 못하였다는 한계점도 존재한다[22]. GAN 모델 연구들에 서 많이 사용되는 데이터 세트는 MNIST, Fashion MNIST, CIFAR-10, Celeb A 등이 있다. 데이터 세트들의 특징은 해상도가 낮으며, 데이터의 복잡도(complexity)가 낮다는 한계를 갖고 있어서 다른 데이터를 사용한 경우 각 모델의 성능을 보장할 수 없다. 따라서 본 논문에서는 의료 영상 합성 연구에서 사용된 모델들을 기준으로 광각 안저사진을 합성하여 결과를 비교한다.

Xin Yi et al(2019)의 연구[23]에 따르면 의료 영상 합성 연구 중 unconditional GAN을 이용한 연구 중에서 DCGAN이 가장 많았으며, 그 뒤로 PGGAN, WGAN 등이 있었다. 그러나 PGGAN의 경우 분할(segmentation)을 목적으로 이미지를 합성하기 때문에 WGAN을 선택하였다. WGAN의 경우 모델이 발표되고 나서 얼마 지나지 않아 WGAN-GP가 발표되어 WGAN 대신 WGAN-GP를 사용하였다. 또한, 기존 연구 중 MRI를 합성하여 알츠하이머의 진단율을 높인 연구 [24]에서 BEGAN 모델을 사용하였기 때문에 이를 선택하여 연구를 진행하였다.

### 1.1. DCGAN [25]

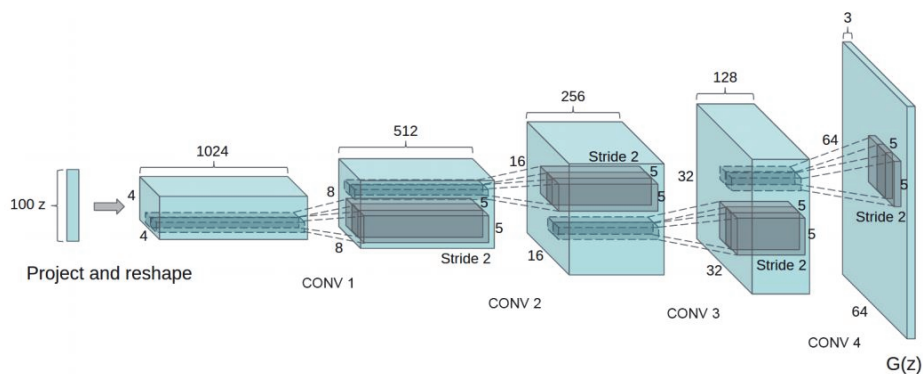
DCGAN(Deep Convolutional GAN)은 2016년 제안되었으며, GAN의 학습을 안정화하기 위해 구조를 변경한 모델이다. 지도학습에서 성공적인 결과를 만든 CNN(Convolutional Neural Networks)을 이용하여 구조를 변경하였다. GAN이

발표된 이후 CNN을 이용해 GAN의 성능을 높이려는 시도는 많았으나, 성공적이지 못하였다. DCGAN 논문에서는 다음의 3가지 특성을 적용하여 성능을 높일 수 있었다고 한다.

생성기의 경우 max-pooling이 아닌 strided convolution을 사용했으며, 판별기의 경우 Up-sampling이 가능하도록 변경하였다. GAN에 사용된 FC(Fully Connected) 레이어가 공간적 정보를 전달하지 못한다는 한계로 데이터의 공간 정보를 전달할 수 있는 컨볼루션(Convolutional) 레이어를 사용하였다. 또한, 각 레이어에 배치 정규화(Batch Normalization)를 이용하였다. 배치 정규화가 제외된 경우는 판별기의 입력 레이어와 생성기의 출력 레이어이다.

3가지 특성 외에도 판별기와 생성기에 적용된 레이어가 다른 특성을 보인다. 판별기의 경우 컨볼루션 레이어를 사용했으며, 활성화 함수는 LeakyReLU를 사용한다. 생성기의 구조는 Deconvolutional layer라고 되어 있으나, 실제 Transpose convolutional layer를 사용하였으며, 활성화 함수로는 ReLU를 사용하였다.

기존의 GAN 모델과 다르게 고해상도 이미지(64\*64)를 생성할 수 있었으며, 대부분 안정적으로 학습시킬 수 있었다. 또한, 생성기에 입력되는 노이즈는 이미지의 특징을 결정하는 것으로, 노이즈 벡터(noise vector)를 저장하여 벡터끼리의 산술 연산을 통해 새로운 이미지가 생성되는 것을 보였다. DCGAN의 구조는 [그림 6]과 같다.

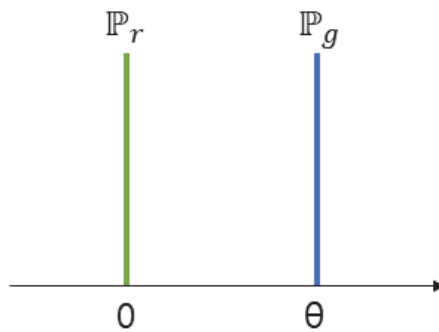


[그림 6] DCGAN 구조 [25]

## 1.2. WGAN[26]과 WGAN-GP[27]

WGAN-GP(Wasserstein GAN - Gradient Penalty)는 WGAN(Wasserstein GAN)이 발표된 이후 WGAN의 단점을 보완하여 발표된 모델이다. WGAN-GP를 이해하기 위해서는 WGAN에 사용된 손실함수를 먼저 알아야 한다. GAN의 성능은 학습이 진행되면서 수렴해야 실제와 같은 데이터를 얻을 수 있다. 이때, 수렴을 판단하는 기준이 두 데이터의 분포상의 거리인데, 기존 모델에 사용된 방법이 가진 한계점을 지적하며, Wasserstein distance를 이용한 모델을 제안한다.

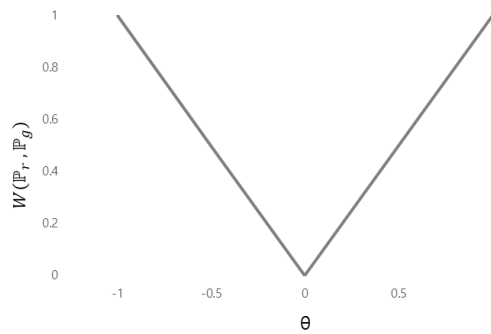
논문에서 비교한 Total Variation(TV) distance, Kullback-Leibler(KL) divergence, Jensen-Shannon(JS) divergence가 가진 한계점을 설명한다. 실제 이미지의 확률분포와 생성된 이미지의 확률분포가 [그림 7]과 같을 때, TV distance의 경우  $\theta$ 가 0이 되지 않는 경우 항상 1을 출력한다. 이러한 문제는 KL divergence와 JS divergence에서도 발견할 수 있다. KL divergence의 경우에는  $\infty$ 값이 나오며, JS divergence의 경우  $\log 2$ 가 출력된다. 이처럼 하나의 값이 계속 출력되는 경우 학습을 진행하지 못하며, 실제 판별기의 학습이 제대로 이루어지지 못하는 이유가 되기도 한다. 이러한 문제를 극복하기 위해 두 확률분포의 차이를 좀 더 유연하게 측정하는 방안이 필요했으며, 해결책으로 Wasserstein distance를 사용하였다.



[그림 7] WGAN - 실제 이미지와 생성된 이미지의 확률분포

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x, y) \sim \gamma} [ \| x - y \| ] \quad (2)$$

Wasserstein distance는 수식 2과 같이 구할 수 있다. 이때, 두 확률분포의 거리를 가장 작게 추정할 값이 출력으로 나오게 되며, 두 확률분포의 거리가  $|\theta|$ 가 된다. 따라서 항상 같은 값을 출력하지 않고,  $\theta$ 값에 따른 두 확률분포의 거리를 나타내면 [그림 8]과 같이 나타낼 수 있다.



[그림 8] WGAN -  $\theta$ 값에 따른 확률분포 거리

WGAN은 Wasserstein distance를 이용해 학습을 안정화했지만, clipping parameter( $c$ )의 값에 따라 판별기가 최적화되기까지 시간이 오래 걸리거나 기울기 소멸(vanishing gradients) 문제가 발생하는 한계점을 갖고 있었다. 모델의 성능이  $c$ 의 값에 따라 좌우되기 때문에 이러한 한계를 극복하기 위한 모델이 WGAN-GP이다. WGAN-GP에서는 clipping을 적용하는 대신 Gradient Penalty를 주는 방식을 택하여 학습의 속도와 안정도를 높였다.

### 1.3. BEGAN [28]

BEGAN(Boundary Equilibrium GAN)은 2017년 구글에서 제안했으며, 모델의 구조가 좀 더 단순하며 빠르고 안정적인 학습이 가능한 모델이다. EBGAN을 바탕으로 오토인코더 구조를 사용한다. GAN은 진짜와 가짜 데이터의 분포를 비슷하게 만들려고 했으나, BEGAN에서는 WGAN의 Wasserstein distance를 손실함수로 이용해 진짜 이미지와 가짜 이미지의 오토인코더 손실(auto-encoder loss)의 분포를 줄이려고 한다.

BEGAN의 특징 중 하나는 diversity ratio( $\gamma$ )를 정의한 것인데, 이는 생성기와

판별기의 균형을 맞추기 위한 하이퍼 파라미터이다. diversity ratio는 0에서 1 사이의 값을 가지며, 가짜 이미지의 품질과 다양성을 조절할 수 있다. diversity ratio에 대한 정의는 수식 3과 같다.

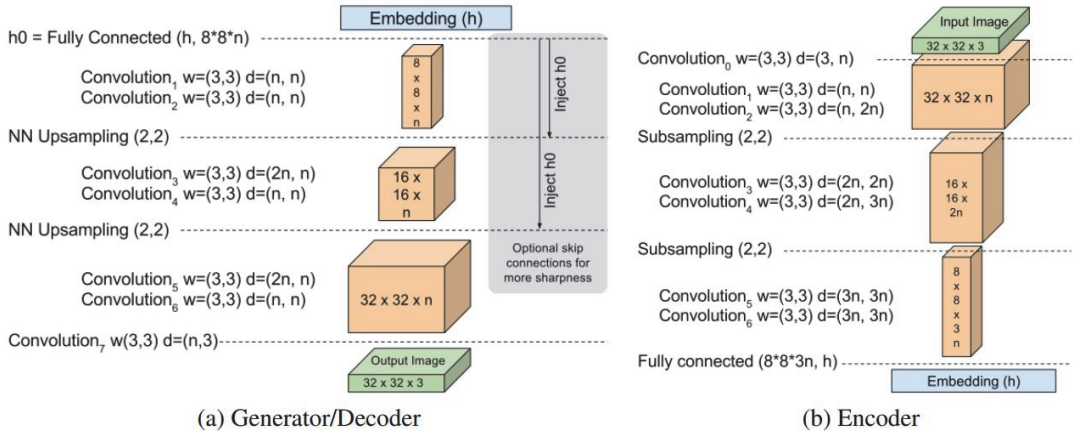
$$\gamma = \frac{E[L(G(z))]}{E[L(x)]} \quad (3)$$

diversity ratio의 값이 작으면 오토인코더인 판별기가 이미지 복원에 집중하기 때문에 생성기에 의해 만들어지는 이미지의 다양성이 떨어지지만, 품질은 높아진다. 반대로 값이 큰 경우 판별기보다 생성기에 좀 더 초점이 맞춰지기 때문에 생성되는 이미지의 다양성이 높아지고, 품질은 떨어지게 된다. diversity ratio 값을 반영한 BEGAN의 목적 함수는 수식 4와 같이 나오게 된다.

$$\begin{cases} L_D = L(x) - k_t L(G(z_D)) & \text{for } \theta_D \\ L_G = L(G(z_G)) & \text{for } \theta_G \\ k_{t+1} = k_t + \lambda_t (\gamma L(x) - L(G(z_G))) & \text{for each training step } k \end{cases} \quad (4)$$

$\theta_D$ 는 판별기의 파라미터를 의미하며,  $\theta_G$ 는 생성기의 파라미터를 의미한다. 첫 번째 식은 판별기의 목적 함수를 의미하며, 생성기의 손실 값을 판별기에 어느 정도 반영할 것인지 결정하여 판별기의 손실 값을 구한다. 두 번째 식은 생성기가 생성한 데이터를 이용해 손실 값을 구한다. 세 번째 식은 Proportional control Theory에서 도입되었으며,  $k$ 의 값에 따라 생성기와 판별기가 서로 견제할 수 있도록 한다. 훈련을 시작하는 경우  $k$ 의 값은 0부터 시작하여, 점차 증가한다.

BEGAN의 경우 DCGAN과 다르게 GAN의 구조를 단순화시켰으며, 배치 정규화, 드롭아웃(Dropout) 등을 사용하지 않는다. BEGAN의 구조는 [그림 9]와 같다.



[그림 9] BEGAN 구조 [28]

## 2. 평가 척도

GAN을 이용해 이미지를 생성하는 경우 평가할 수 있는 척도는 IS(Inception Score), FID(Frechet Inception Distance)가 있으며, 이 외에는 GAN 판별기의 분류 정도를 이용하거나 이미지 품질 평가 방법을 이용한다. 이미지 품질 평가 방법에는 SSIM, MSE, MAE, PSNR 방법이 대표적이다.

IS와 FID 모두 ImageNet의 데이터로 미리 훈련된 Inception 모델을 사용하는 공통점이 존재한다. IS의 경우 생성된 이미지를 Inception 모델에 넣어 얻을 수 있는 확률 벡터를 이용하여 점수를 계산한다. 해당 수식은 수식 5와 같다.

$$IS(G) = \exp(E_{x \sim p_a} D_{KL}(p(y|x) \| p(y))) \quad (5)$$

IS는 이미지 생성에 사용된 실제 이미지의 확률 분포를 반영하지 못한다는 단점이 존재한다. 실제 데이터의 분포를 반영하기 위해서 나온 척도이며, FID는 수식 6과 같이 구할 수 있다.

$$FID(x, g) = \| \mu_x - \mu_g \|_2^2 + Tr\left(\sum^x + \sum^g - 2\left(\sum^x \sum^g\right)^{\frac{1}{2}}\right) \quad (6)$$

FID는 실제 이미지와 생성된 이미지를 모두 Inception 모델에 넣고 각각의 확률 벡터를 이용해 평균과 공분산 값을 이용해 점수를 측정한다. IS의 경우 값이 클수록 좋은 결과이며, FID의 경우 값이 작을수록 좋은 결과에 해당한다.

이미지 품질 평가에는 MSE(Mean Squared Error), MAE(Mean Absolute Error), PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity) 방법이 존재한다. 각 방법은 픽셀값을 이용하여 이미지를 비교한다.

MSE와 MAE는 이미지 평가에도 사용되지만, 회귀에서도 자주 사용되는 손실 함수이다. MSE는 오차의 제곱에 대한 평균으로 값이 작은 경우 실제와 비슷하다는 것을 나타낸다. MAE는 실제 데이터와 생성된 데이터의 절대 오차의 평균으로 값이 작은 경우가 실제와 비슷한 이미지를 생성한 것이다. 각 값을 구하는 식은 수식 7, 8과 같다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x| \quad (8)$$

PSNR은 이미지 정보와 화질을 평가하는 경우 사용되는 척도로 생성된 이미지의 품질이 좋은 경우 값이 커지는 특징을 보인다. 해당 값은 수식 9와 같이 구할 수 있으며,  $MAX_I$ 는 해당 이미지의 픽셀 최댓값과 최솟값의 차이이다. 이미지는 RGB 채널로 이루어져 있으므로 픽셀의 최댓값은 255이고 최솟값은 0이기 때문에  $MAX_I$  값은 255가 된다.

$$PSNR = 20 \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (9)$$

PSNR을 사용하는 경우 실제 사람이 느끼는 이미지의 품질과 다른 경우가 존재한다. 사람은 구조적인 부분을 포함하여 이미지를 보지만 PSNR은 픽셀값을 이용하여 계산하기 때문에 이미지의 구조적인 부분을 포함하지 못하는 단점이 존재한다. 이러한 한계를 극복하기 위해 등장한 방법이 SSIM이다. SSIM은 해당 이미지가 구조적으로 얼마나 일치하는지 평가할 수 있는 척도이다. SSIM은 이미지의 휘도(Luminance), 대비(Contrast), 구조(Structure)를 계산하여 SSIM 점수를 측정한다. 실제 이미지를  $x$ , 생성 이미지를  $y$ 라 할 때, SSIM 점수는 수식 10과 같이 구할 수 있다.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (10)$$

본 연구에서는 위에서 설명한 평가척도들을 이용하여 생성된 이미지의 정량적 평가를 진행하고 GAN의 생성 이미지 품질 평가를 진행하려고 한다.

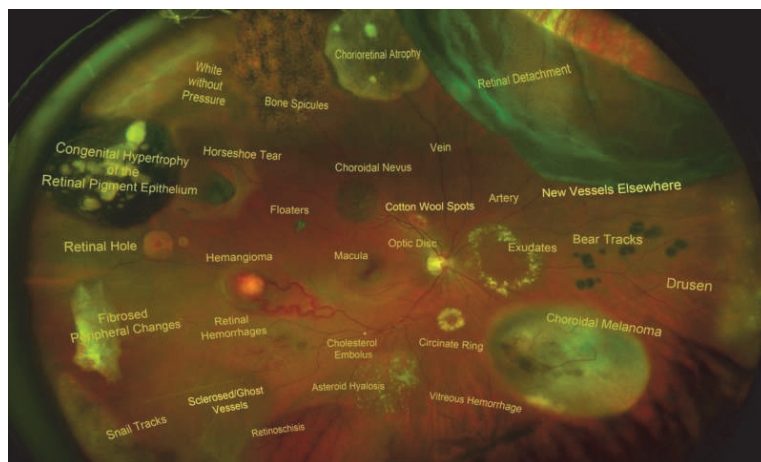


## IV. 합성 모델 설계 및 결과 비교

이 장에서는 본 연구 진행 방법 및 결과를 살펴본다. 첫 번째로 연구에 사용된 실험 데이터에 대해 기술하였다. 실험 데이터의 수집, 라벨링, 전처리 과정을 살펴본다. 두 번째로는 본 연구에서 사용된 세 모델 구성 등을 살펴보고, 실험 환경을 기술하였다. 세 번째로는 세 모델이 학습하며 변하는 과정을 합성 결과를 통해 살펴보았다. 학습 단계를 초기, 중기, 말기로 나눠 단계별 합성 결과를 커널 밀도 추정을 이용하여 데이터 분포의 변화를 살펴보았다. 마지막으로 실험 결과를 비교하기 위해 정성적 평가와 정량적 평가로 나누어 제시하였다.

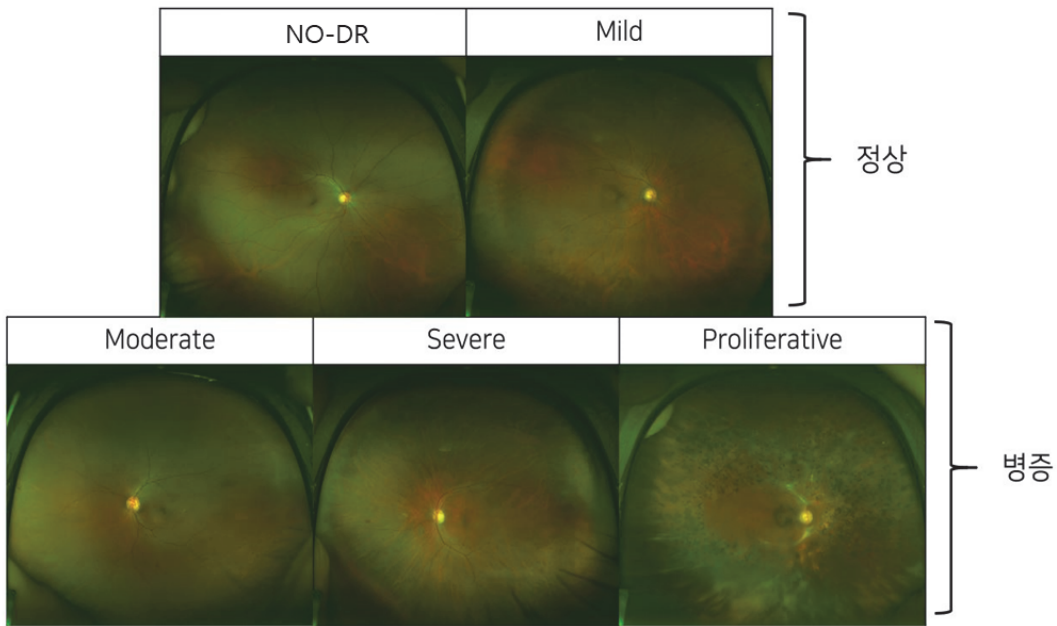
### 1. 실험 데이터

본 실험에서 사용된 데이터는 제주대학병원 안과와 협업을 통해 총 152장의 광각 안저사진을 수집하였다. 사용된 GAN 모델이 Unconditional GAN으로 데이터의 라벨링이 필요하지 않다. 그러나 광각 안저사진을 통해 검출될 수 있는 질병이 다양하며, 그 증상 또한 다양하므로 당뇨망막병증으로 한정하여 데이터를 수집하였다. [그림 10]은 광각 안저사진에서 검출될 수 있는 증상을 보여준다.



[그림 10] 광각 안저사진에서 검출 가능한 증상 (출처 : Optos)

데이터는 국제 임상 당뇨병망막증(ICDR: International Clinical Diabetic Retinopathy) 중증도의 기준에 따라 No-DR, Mild, Moderate, Severe, PDR로 라벨링 되었다. 5개의 등급 중 Severe 등급이 32장, 이외 등급에서는 각각 30장의 데이터를 수집하였다. 해당 데이터는 개인 정보가 삭제된 상태에서 받았으며, 각각 파일명은 '등급(순번)'으로 되어있다. 등급별 광각 안저사진은 [그림 11]과 같다.



[그림 11] 당뇨병망막병증 중증도에 따른 안저사진 (출처 : 제주대학병원 안과)

데이터가 5가지 세트로 나누어져 있고, 세트별로 존재하는 데이터의 수량이 적다. 또한, 중증도에 따라 관측되는 현상이 달라서 데이터를 재구성하여 본 실험에 적용하였다. 정상과 병증이 존재하는 등급으로 나뉘었으며, 정상은 실제 No-DR, Mild 등급에 해당하며, 병증의 경우 Moderate, Severe, PDR 등급에 해당한다. 그 결과 정상의 경우 60장, 병증의 경우 92장을 사용하여 실험을 진행하였다.

광각 안저사진의 원본 사이즈는 3900\*3072이며, 실험에 적용하기 위해 사이즈를 줄여 사용하였다. 기존의 광각 안저사진을 이용한 연구에서 96\*96 또는 256\*192, 512\*512 사이즈를 이용했기 때문에 사진 비율을 유지하며 실험 가능한 사이즈인 256\*192를 사용하였다.

본 실험을 위해 데이터의 전처리는 따로 진행하지 않았다. 일부 사진에서 관찰자

의 손, 환자의 속눈썹, 빛 반사 등 노이즈를 발견할 수 있었으나, 이러한 부분을 제거하기 위해서는 모든 사진을 같은 사이즈로 잘라야 한다. 그러나 광각 안저사진은 일반 안저사진에 비해 관찰할 수 있는 범위가 넓다는 것이 장점인데, 노이즈 제거를 위해 사진을 자르는 경우 그 장점을 상쇄시킬 수 있으므로 이미지를 잘라내는 작업은 진행하지 않았다.

## 2. GAN 모델 및 실험 환경

GAN은 각 모델의 특징이 달라 같은 환경에서 학습시키기 어려운 점이 존재한다. 따라서 각 모델의 구현은 모델이 발표된 논문을 기반으로 하였으나, 기존 연구에서 사용된 이미지보다 해상도가 높아서 레이어를 추가하였다.

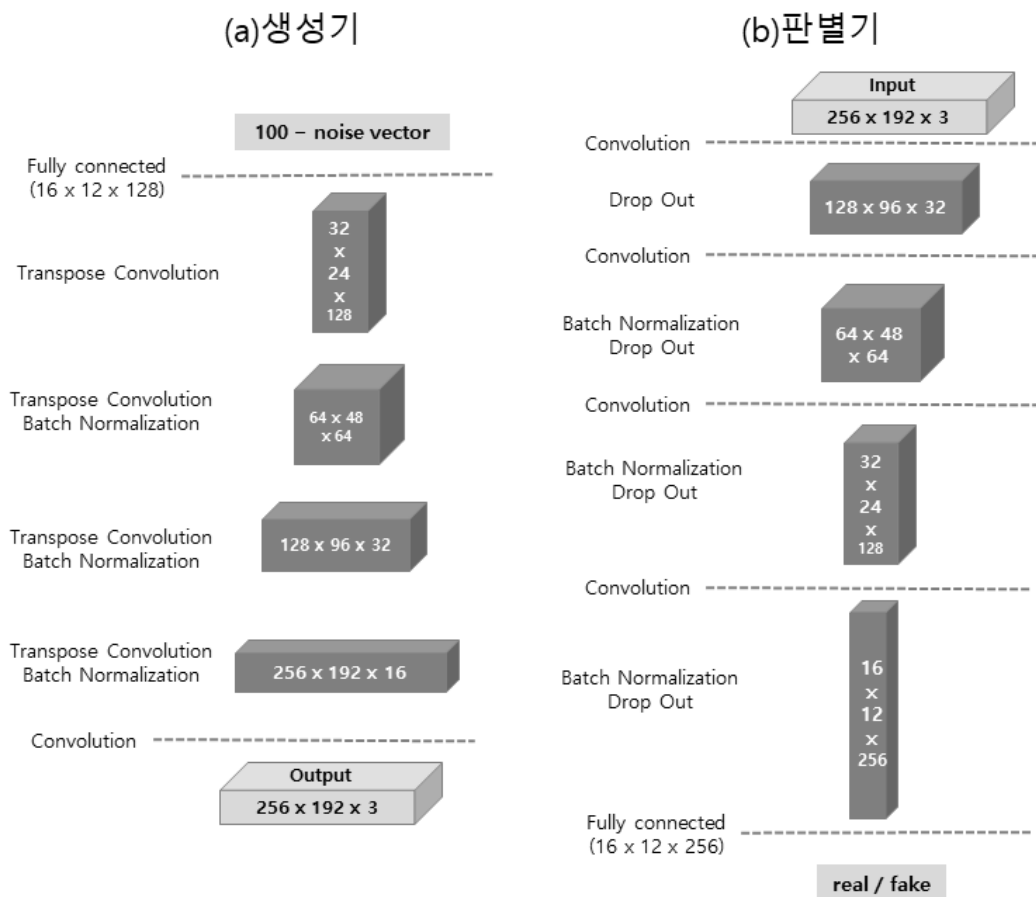
GAN의 하이퍼 파라미터는 배치 정규화, 드롭아웃, 학습률, 활성화 층, 합성곱 필터, 커널 크기, 스트라이드, 배치 크기, 잠재 공간 크기 등 조절해야 하는 값이 다른 신경망에 비해 많다. 각 모델의 하이퍼 파라미터가 다르므로 같은 값을 설정할 수 있는 일부 하이퍼 파라미터를 선택하였다. 같은 값을 갖는 하이퍼 파라미터는 잠재 공간의 사이즈, 생성된 이미지의 사이즈, 배치 사이즈를 같게 설정하였다. 잠재 공간의 사이즈는 100이고, 배치 사이즈는 8로 설정하였고, 생성된 이미지의 사이즈는 광각 안저사진 연구에서 많이 사용되었던 256 \* 192 사이즈를 선택하였다.

모델의 구현은 TensorFlow 1.14와 Keras 2.2를 이용하였으며, 개발 환경은 <표 3>과 같다.

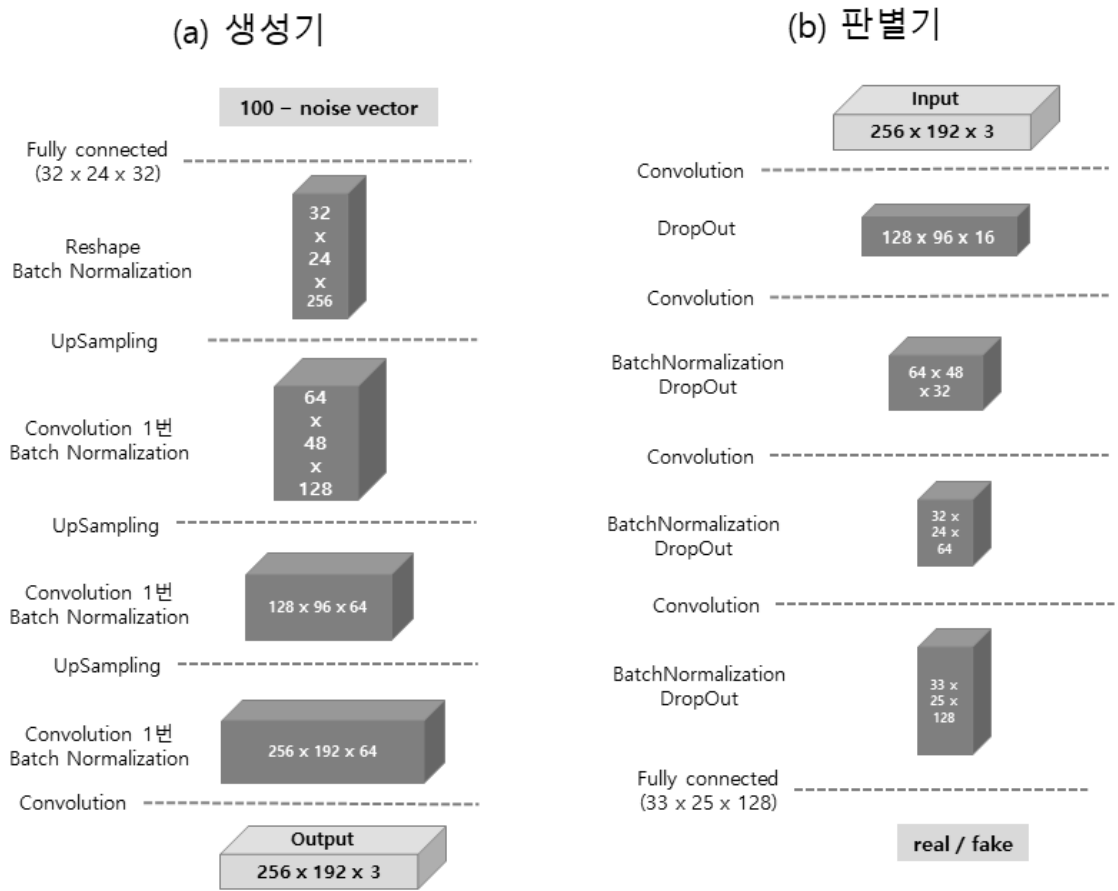
<표 3> 개발 환경

항목	사양
운영체제	Ubuntu 18.04 LTS
CPU	IntelCore®Processor i7-7700 3.60Ghz
SSD	ADATA SU650 (240GB)
RAM	16GB DDR4
GPU	NVIDIA Geforce GTX1060Ti

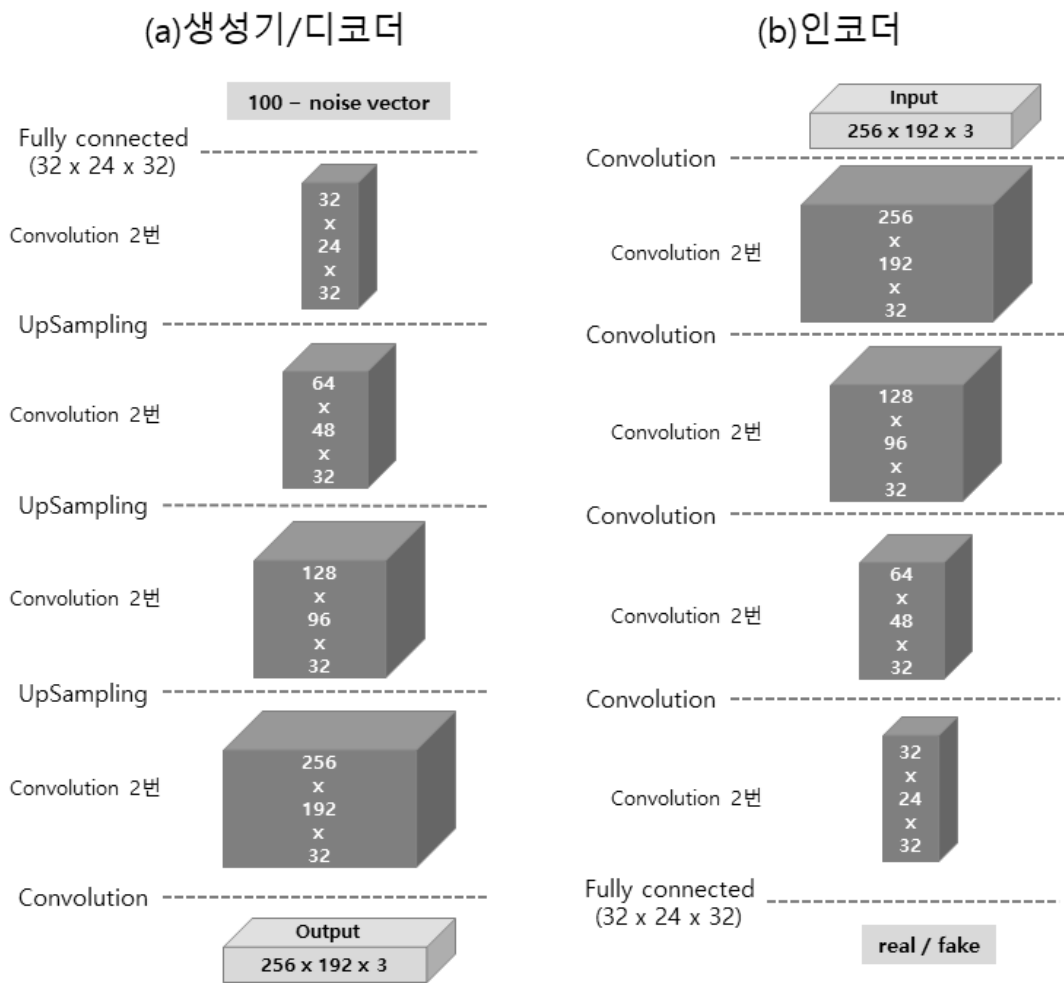
세 모델의 구현은 발표된 논문을 기준으로 하였으나 본 실험에서 사용되는 데이터의 사이즈가 크기 때문에 일부 레이어를 추가하여 사용하였다. 레이어를 추가하는 경우 기존 논문에서 사용된 레이어의 하이퍼 파라미터(커널 크기, 스트라이드 등)를 유지하면서 추가하였다. DCGAN의 생성기의 경우 Transpose Convolutional layer가 하나 더 추가되었다. WGAN-GP의 경우 생성기에 Up-sampling과 Convolution 연산이 추가되었으며, BEGAN의 경우 디코더에는 Up-sampling과 Convolution 2번 연산이 추가되었으며 인코더의 경우 Convolution 연산이 3번 추가되었다. 이외의 세팅은 발표된 논문과 같게 하였으며, 구현된 각 모델의 구조는 [그림 12] ~ [그림 14]에서 확인할 수 있다.



[그림 12] 구현한 DCGAN 구조



[그림 13] 구현한 WGAN-GP 구조

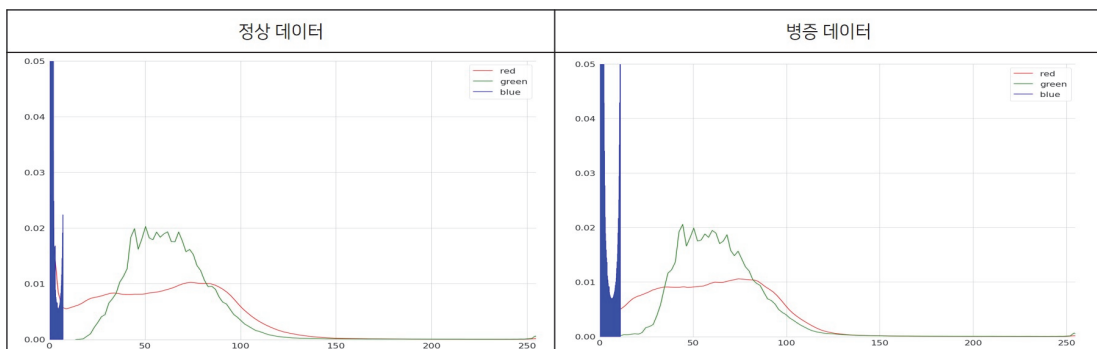


[그림 14] 구현한 BEGAN 구조

### 3. 학습 단계별 데이터 분포 변화

각 모델이 학습하는 과정에서 데이터 분포가 어떻게 변하는지를 비교하기 위해 커널 밀도 추정(Kernel Density Estimation)을 이용해 데이터 분포를 나타내었다. 커널 밀도 추정은 히스토그램을 커널함수(Kernel Function)를 이용해 확률밀도함수로 만드는 방법이다. 히스토그램의 경우 계급 간에 불연속성이 나타난다는 특징이 있고, 계급의 크기에 따라서 히스토그램이 달라지는 단점이 존재한다. 따라서 이미지의 각 R, G, B값을 커널 밀도 추정을 통해 데이터의 분포를 비교하고자 한다. B값의 경우 0 ~ 10계급 사이에 분포하고 있는 데이터의 양이 많아 실제 밀도 값이 다른 값에 비해 증가한 경우가 존재했다. 각 값의 정확한 비교를 위해 그래프의 세로축인 밀도 값을 0.05로 한정 지어 출력하였다.

훈련 과정을 20,000 epoch를 기준으로 100, 10,000, 20,000 epoch를 초기, 중기, 말기로 나누어 나타냈으며 DCGAN의 경우 학습 수렴 속도가 느려 학습 초기를 1,000 epoch를 기준으로 데이터 분포를 나타내었다. 실제 데이터의 분포는 [그림 15]에 해당하며, 각 모델의 이미지 분포는 [그림 16] ~ [그림 25]와 같다.

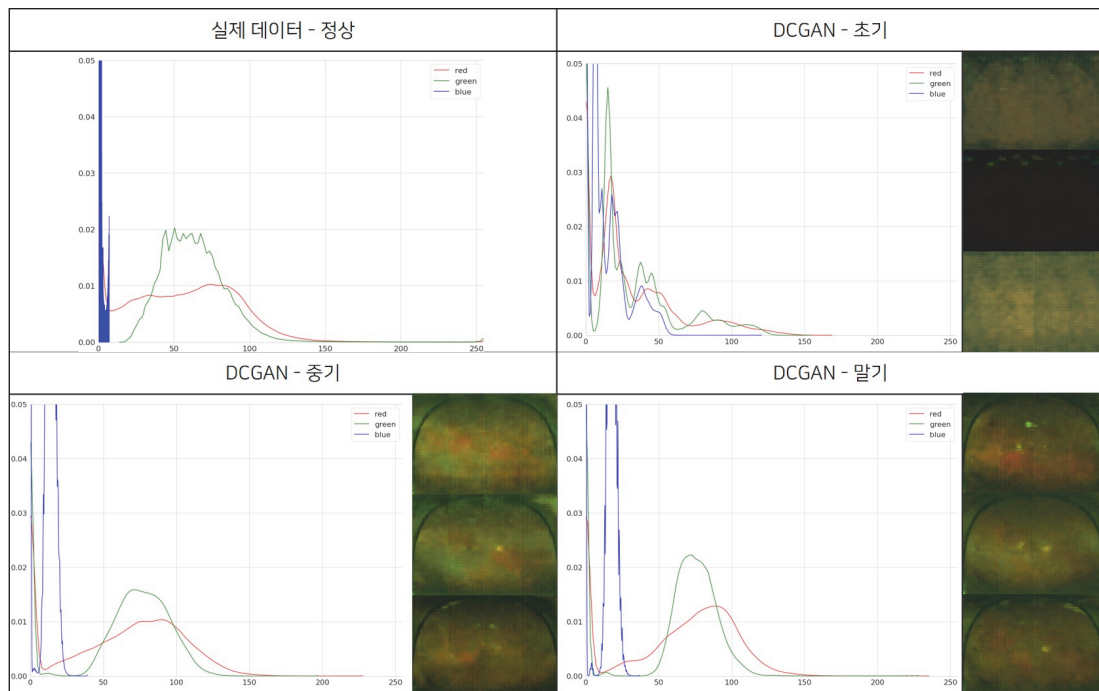


[그림 15] 정상 데이터와 병증 데이터의 분포

실제 데이터의 분포를 비교하면 정상과 병증의 데이터 분포가 유사하다는 것을 알 수 있다. 실험 결과 부분에서 활용한 평가 척도에 따르면 똑같은 데이터 세트라도 특징이 다르다는 것을 알 수 있으나, 실제 색의 분포를 보면 유사한 것을 알 수 있다. 초록색의 경우 이미지 분포가 유사하나 빨간색과 파란색의 분포가 눈에 띄게 다른 것을 보면 병증의 유무를 결정하는 색이 빨간색과 파란색이라는 것을

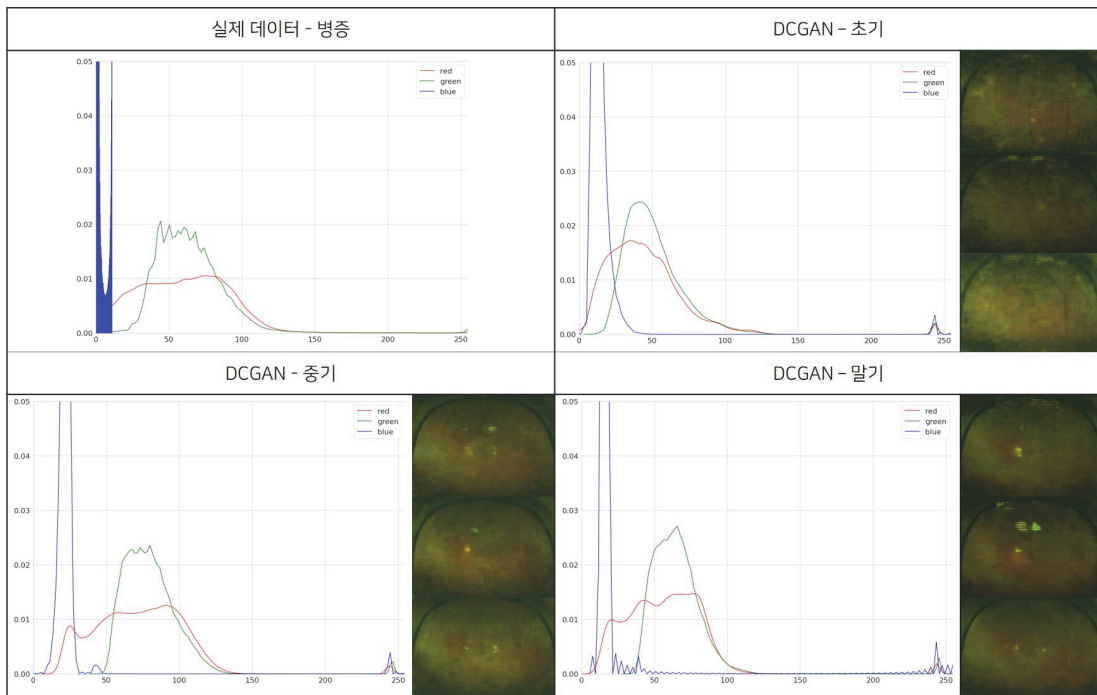
알 수 있다. 따라서 빨간색과 파란색의 분포가 실제 데이터와 유사한 경우 병증 구분할 수 있을 것이라 기대할 수 있다.

DCGAN([그림 16, 17])의 학습 과정을 살펴보면 데이터의 분포 모습이 실제 데이터와 비슷해지는 것을 알 수 있으나 파란색이 실제 데이터보다 부족하다는 것을 알 수 있다. 또한, 병증이 있는 데이터를 생성하는 경우 파란색 데이터 분포가 실제 데이터에서는 존재하지 않았던 250에 가까운 값들의 분포가 생긴 것을 알 수 있다. 병증이 존재하지 않는 이미지를 합성한 경우보다 병증이 있는 이미지를 합성한 경우 결과물이 어두워 보이는 것이 해당 데이터의 분포가 영향을 미쳤다는 것을 알 수 있다.



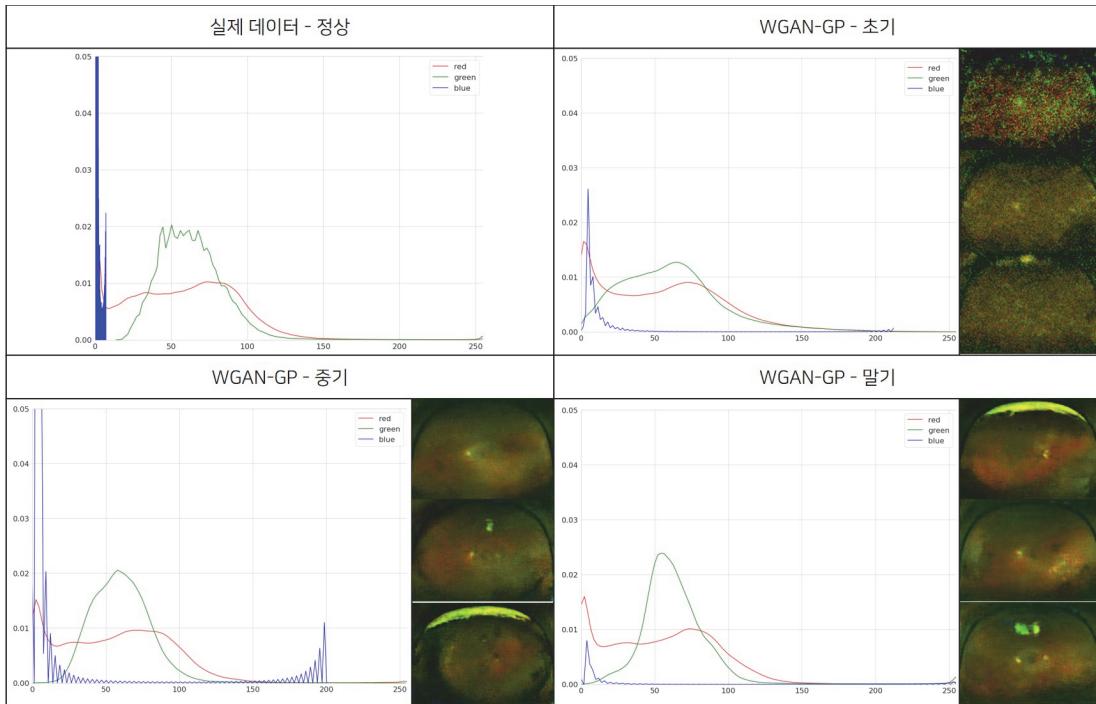
[그림 16] DCGAN 학습 과정 - 정상 데이터 사용



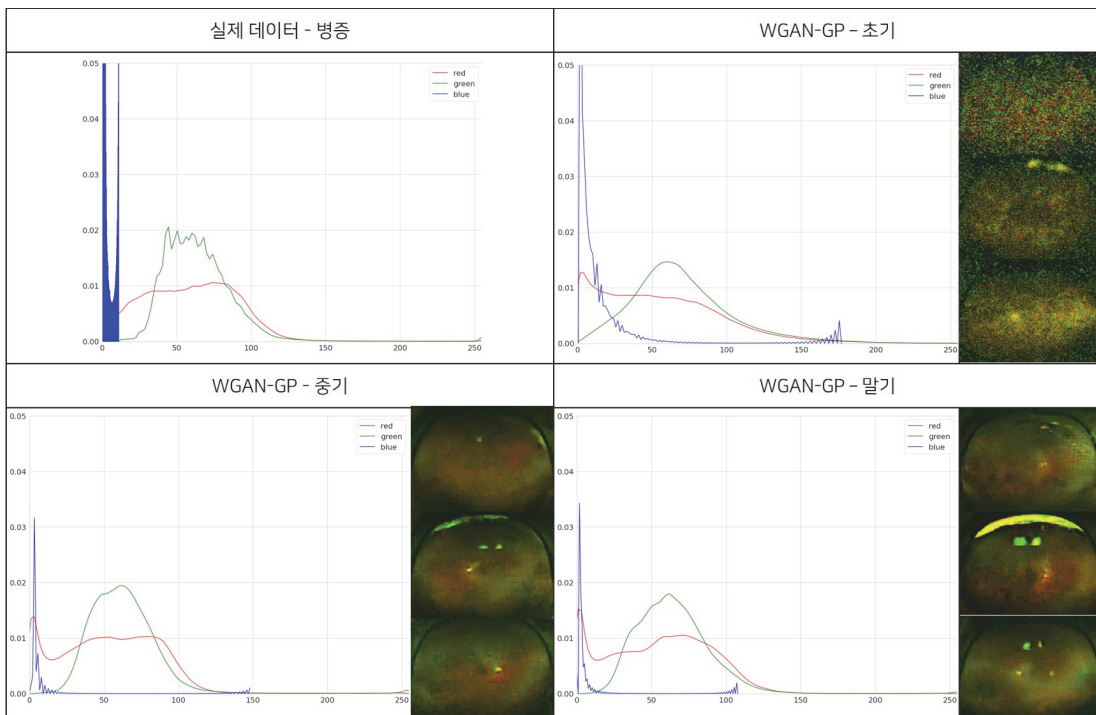


[그림 17] DCGAN 학습 과정 - 병증 데이터 사용

WGAN-GP의 경우([그림 18, 19]) 세부적인 그래프 모형은 다르나 실제 데이터와 비교할 때 값의 범위가 가장 유사한 것을 알 수 있다. 이러한 점은 실제 WGAN-GP가 Wasserstein distance를 목적함수로 사용하기 때문이라 추측할 수 있다. 그러나 세 모델 중 파란색 분포의 밀도가 가장 떨어지며, 각 그래프가 완만한 모양을 보이는 것을 알 수 있다. 빨간색의 경우 실제 데이터와 매우 유사한 분포를 보인다.



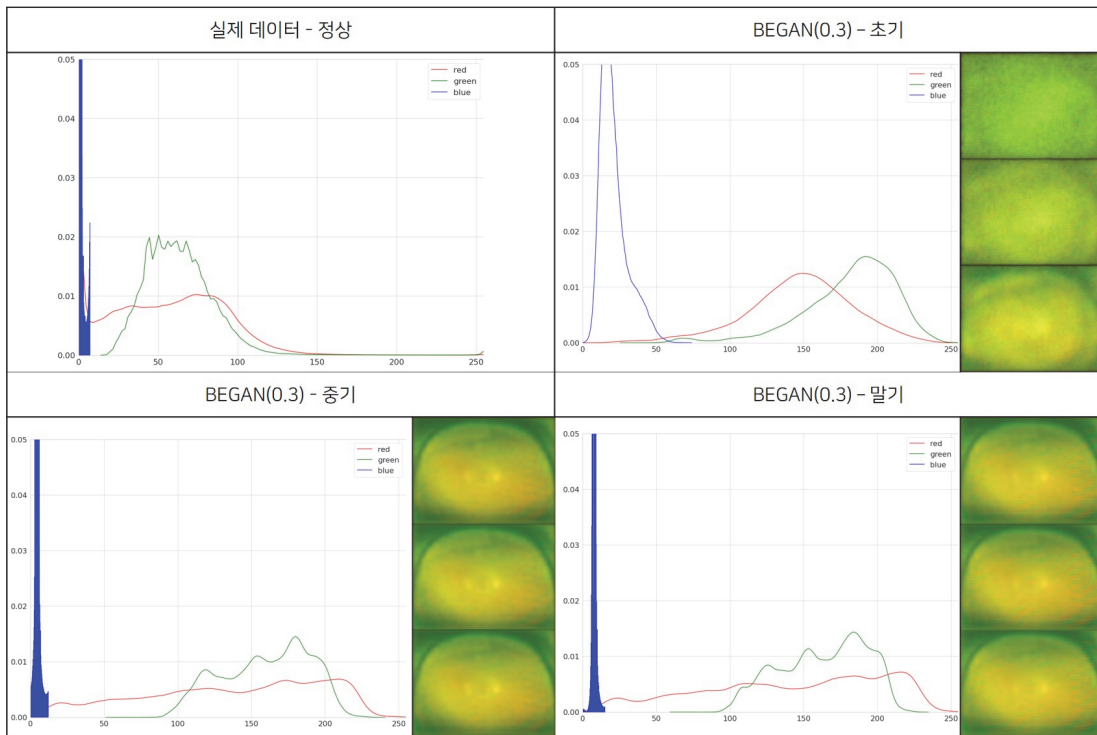
[그림 18] WGAN-GP 학습 과정 - 정상 데이터 사용



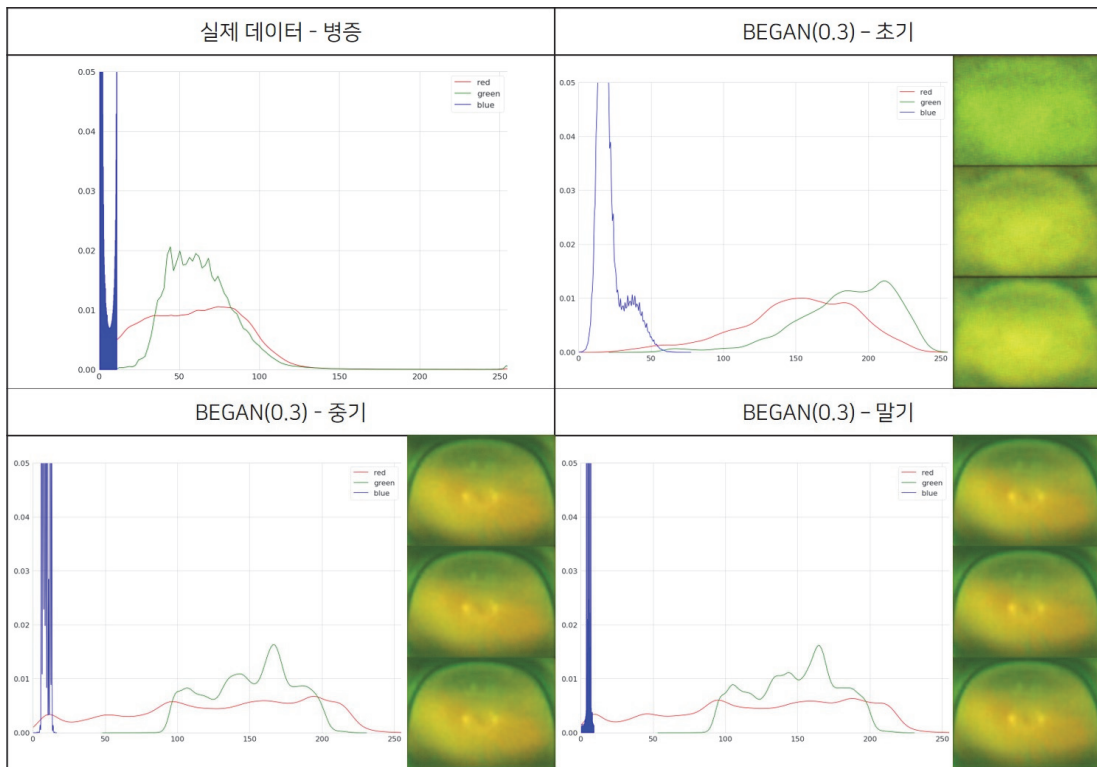
[그림 19] WGAN-GP 학습 과정 - 병증 데이터 사용

BEGAN의 경우([그림 20~25]) 세 모델 중 가장 파란색 밀도가 유사한 모델이었다. 각 학습 과정을 살펴보면 diversity ratio 값이 증가함에 따라 데이터 분포의 거리가 가까워지는 것을 알 수 있다. 그러나 WGAN-GP와 비교하면 BEGAN의 결과 분포 거리가 좀 더 먼 것을 알 수 있다. 이러한 차이가 발생하는 이유는 BEGAN의 경우 실제 데이터 분포가 아닌 오토 인코더를 통해 도출된 복원된 이미지 데이터 분포를 이용하기 때문이다. 따라서 BEGAN의 결과 이미지의 색 밀도 정도가 실제 이미지보다 떨어지는 것을 알 수 있다.

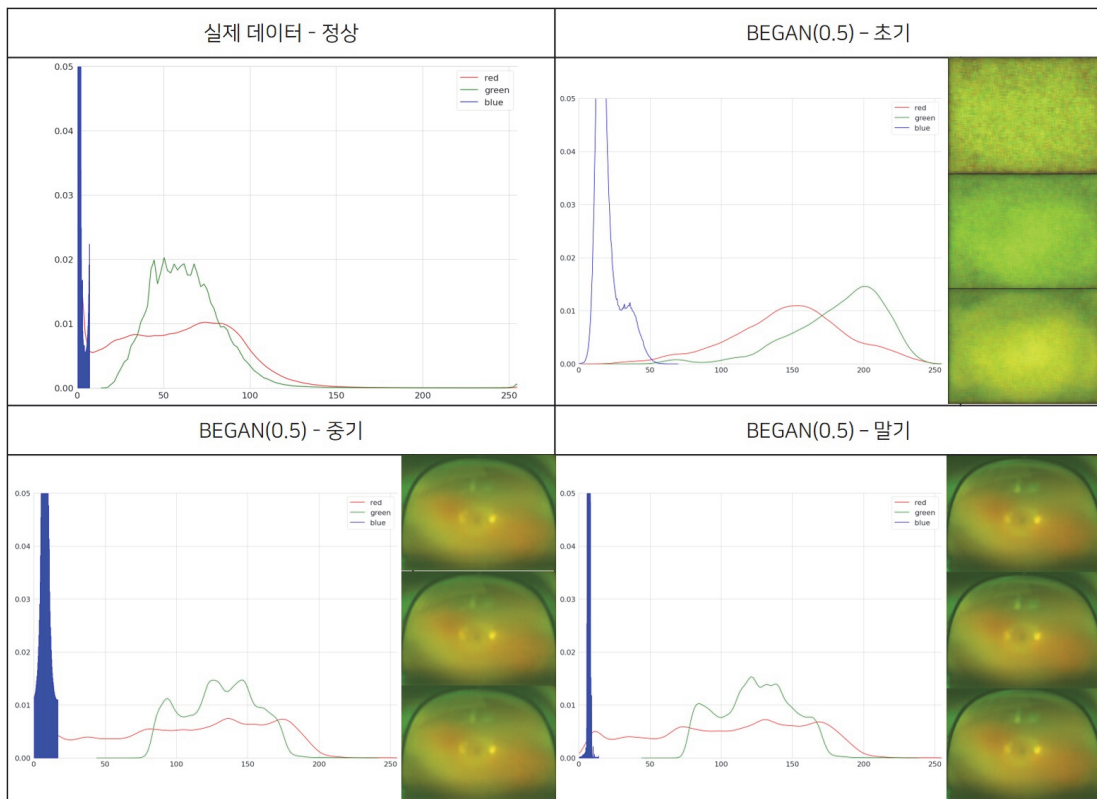
diversity ratio가 0.3이고 사용된 데이터가 정상인 경우를 제외하고서는 빨간색 데이터의 분포가 완만하게 퍼지는 것을 알 수 있다. 그러나 실제 데이터에서는 빨간색이 완만하지 않고 100 미만의 값에 쏠려 있는 것을 알 수 있다. 또한 파란색 데이터가 실제와 비슷한 밀도를 보여 생성된 이미지의 밝기가 세 모델 중 가장 밝은 결과를 도출하였다.



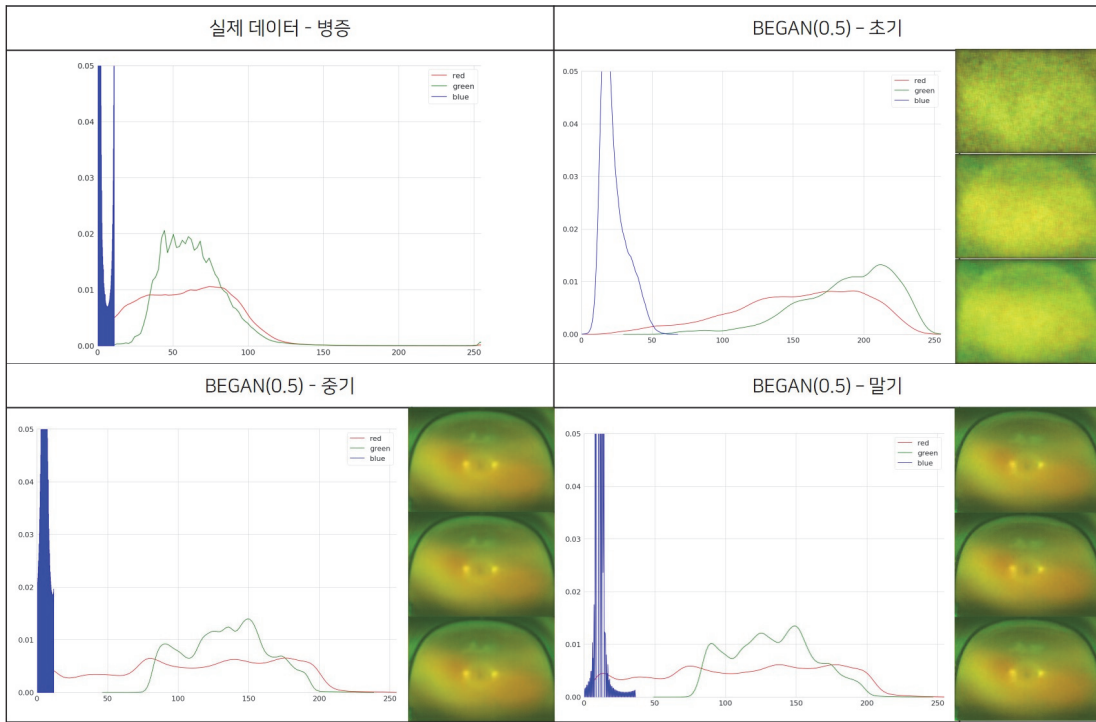
[그림 20] BEGAN(0.3) 학습 과정 - 정상 데이터 사용



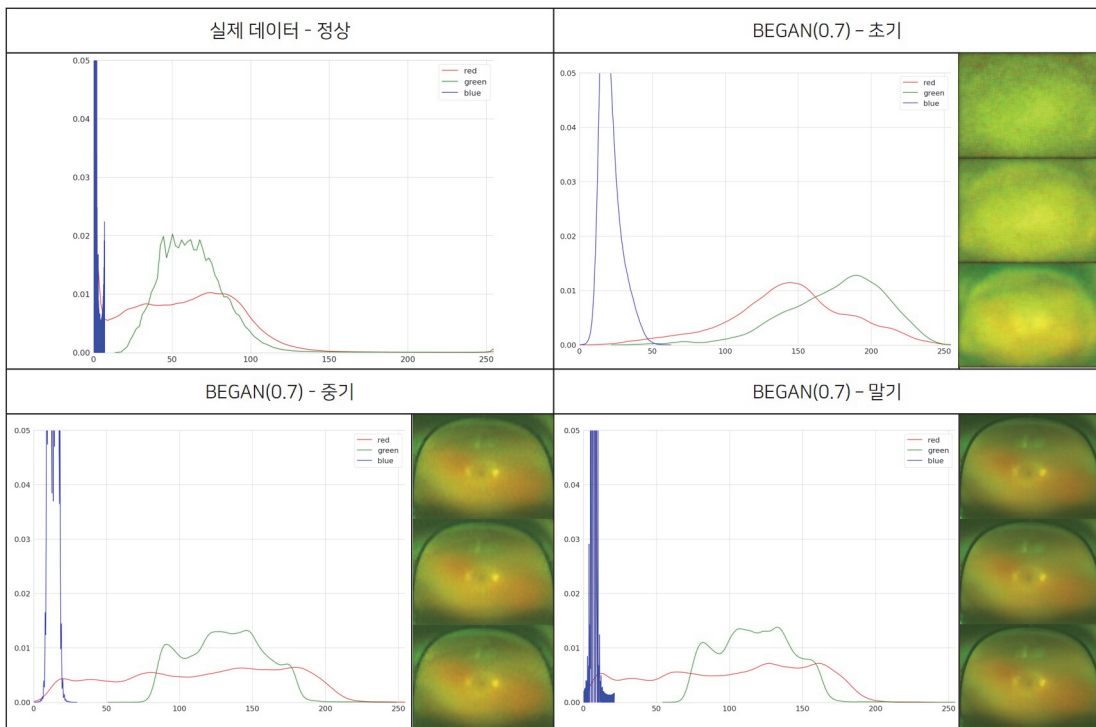
[그림 21] BEGAN(0.3) 학습 과정 - 병증 데이터 사용



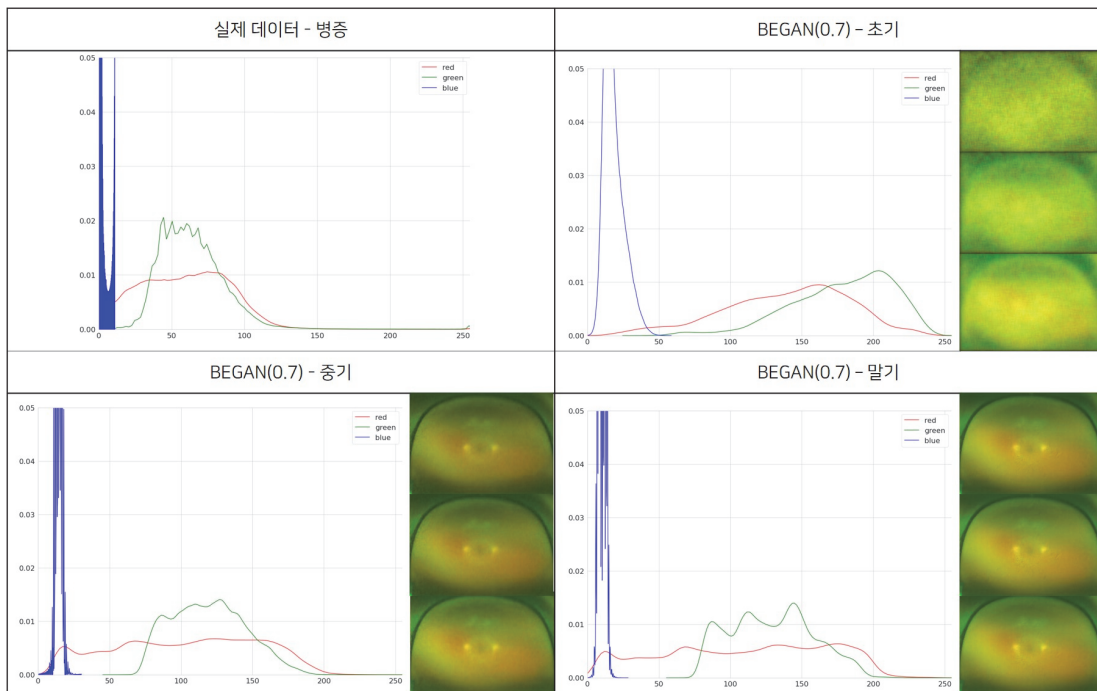
[그림 22] BEGAN(0.5) 학습 과정 - 정상 데이터 사용



[그림 23] BEGAN(0.5) 학습 과정 - 병증 데이터 사용



[그림 24] BEGAN(0.7) 학습 과정 - 정상 데이터 사용



[그림 25] BEGAN(0.7) 학습 과정 - 병증 데이터 사용

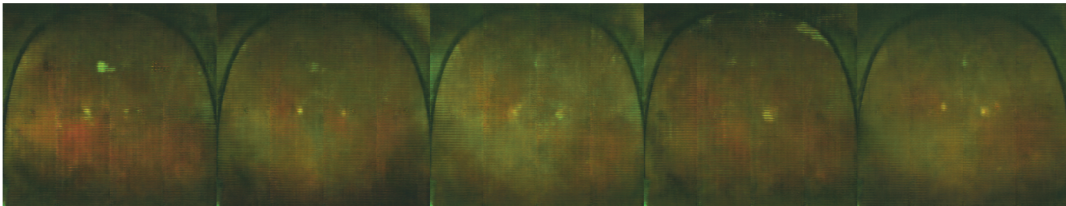
이미지의 색 분포를 살펴보면 세 모델 모두 다르게 변하고 있는 것을 알 수 있으며, 각 색의 값이 이미지에 어떤 영향을 미치는지 알 수 있다. 따라서 데이터 합성 과정에서 각 이미지의 색 분포가 어떤 영향을 미치는지 알 수 있다면 이미지를 생성하는 과정에서 각 색의 값의 중요도를 다르게 할 수 있다. 색의 중요도를 다르게 한다면 GAN의 새로운 하이퍼 파라미터가 될 수 있을 것이다. 또한, 색의 중요도를 모델마다 다르게 한다면 실제 이미지 분포와 유사한 이미지를 생성할 수 있을 것으로 기대된다. BEGAN의 경우 파란색이 가장 비슷하였으며, WGAN-GP의 경우 빨간색이 가장 비슷한 분포를 보였다. 이미지를 구성하기 위해서는 R, G, B 값이 모두 비슷해야 하므로 학습 과정에서 비슷한 양상을 보이지 못하는 색의 값에 중요도를 추가한다면 실제 데이터와 유사한 결과를 얻을 수 있을 것으로 기대된다.

#### 4. 합성 결과 비교

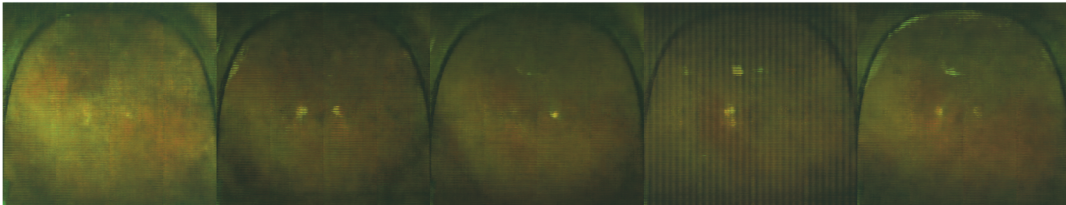
모델별 실험 결과는 정성적 평가와 정량적 평가로 나누어 진행하였다. 정성적 평가의 경우 사진에서 관찰할 수 있는 요소들을 나눠 정리하였으며, 모델별 실험 결과를 제시하였다. 정량적 평가의 경우 FID를 측정하였고, 비교 가능한 기준값이 존재하지 않아 실제 훈련 데이터 세트의 FID 값을 기준으로 비교하였다.

##### 4.1. 정성적 평가

모델별 안저사진의 합성 결과는 [그림 26] ~ [그림 28]과 같다. 모델별 정상과 병증 데이터 세트의 합성 결과를 제시하였으며, 20,000 epoch를 기준으로 실험을 진행하였다. BEGAN의 경우 diversity ratio( $\gamma$ ) 값을 0.3, 0.5, 0.7로 실험을 진행하였기 때문에 각 경우를 나누어 결과를 제시하였다.

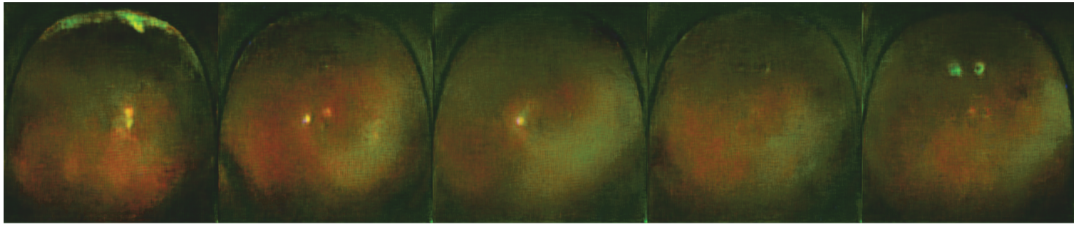


(a) 정상안 데이터를 이용한 합성 결과

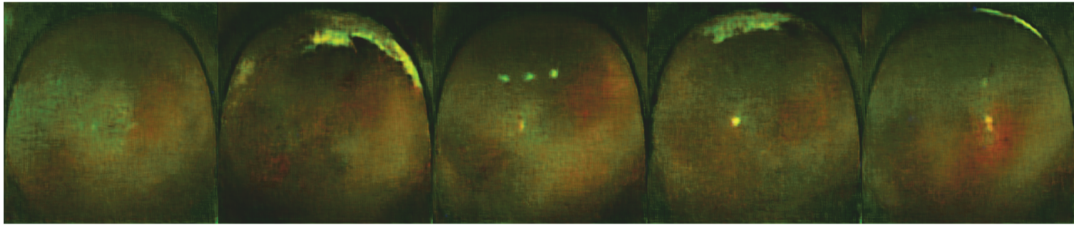


(b) 증상이 있는 데이터를 이용한 합성 결과

[그림 26] DCGAN 합성 결과 (20,000 epoch)

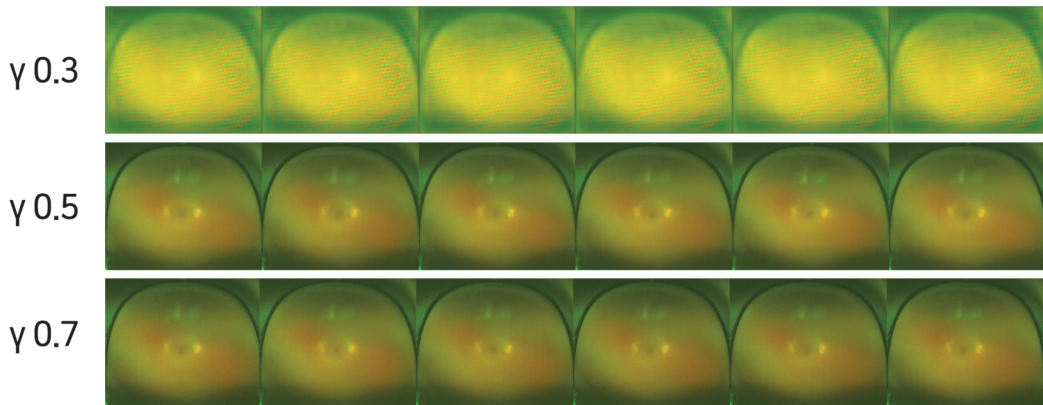


(a) 정상안 데이터를 이용한 합성 결과

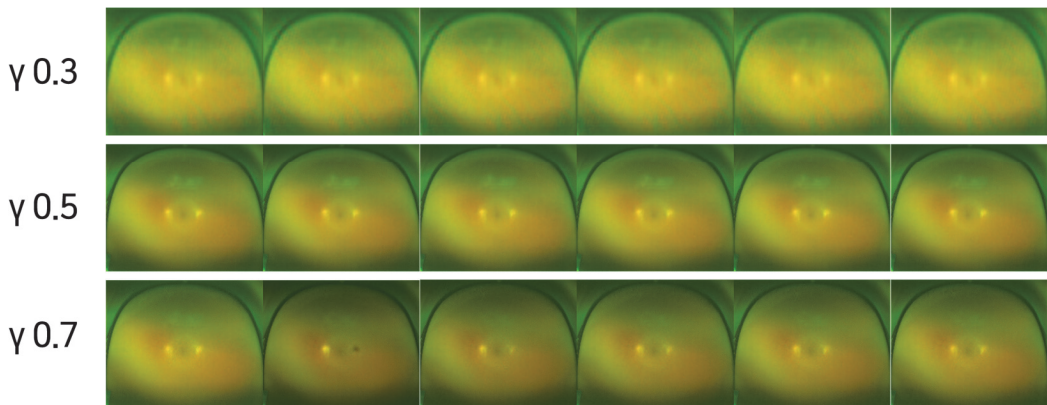


(b) 증상이 있는 데이터를 이용한 합성 결과

[그림 27] WGAN-GP 합성 결과(20,000 epoch)



(a) 정상안 데이터를 이용한 합성 결과



(b) 증상이 있는 데이터를 이용한 합성 결과

[그림 28] BEGAN 합성 결과 (20,000 epoch)



안저사진에서 관찰되어야 하는 요소들은 시신경유두(Optic disk), 황반, 혈관이 있으며 추가로 정상 데이터와 병증 데이터를 구별할 수 있어야 하고, 망막의 윤곽을 식별할 수 있어야 한다. 따라서 데이터의 평가를 5가지 요소로 결정하여 평가하였으며, 모든 사진에서 관찰할 수 있는 경우 ○, 일부 사진에서 관찰할 수 있는 경우 △, 관찰할 수 없는 경우 ×로 표시하였다. 결과는 <표 4>와 같다.

<표 4> 모델별 안저사진 요소 평가

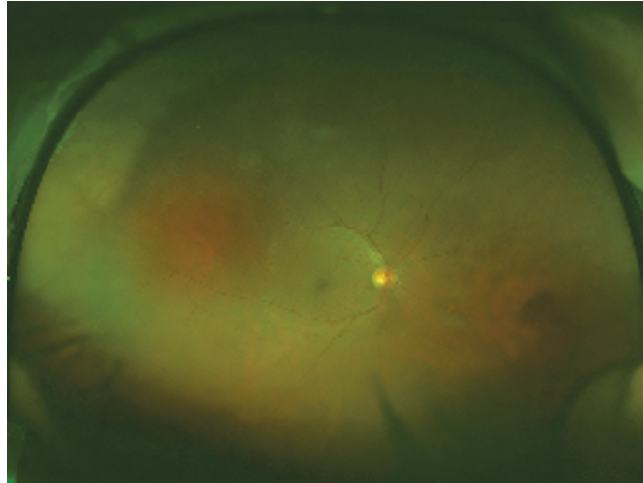
구분기준	DCGAN	WGAN-GP	BEGAN(0.7)
윤곽	○	○	○
시신경유두	△	△	△
황반	×	△	○
혈관	×	×	×
정상/병증 구분	×	×	△

DCGAN의 경우 전체적인 윤곽을 확인할 수 있었으나, 시신경유두의 경우 일부 사진에서는 관찰할 수 있지만, 일부에서는 존재하지 않았다. 또한, 경계가 모호해 황반의 구별이 쉽지 않았으며, 혈관과 정상, 병증 구별이 어려웠다.

WGAN-GP의 경우 윤곽뿐만 아니라 눈꺼풀 등 기존 사진에 존재하는 노이즈도 함께 관측할 수 있었다. 그러나 시신경유두와 황반의 경우에는 일부 사진에서만 관찰할 수 있었다. DCGAN과 마찬가지로 혈관을 관찰하기 어려웠고, 병증과 정상안 구별이 힘들었다.

BEGAN의 경우 diversity ratio 값에 따라 생성된 결과가 조금 다르지만 0.7의 값의 결과를 기준으로 보면 전체적인 윤곽은 뚜렷하게 확인할 수 있었다. 그러나 시신경유두의 경우 1개가 관찰되어야 하는데 2개가 관찰되었으며, 황반의 경우 생성된 모든 이미지에서 확인할 수 있었다. 혈관은 관찰할 수 없었으나 정상과 병증이 존재하는 경우 이미지가 다른 것을 알 수 있었다.

각 모델이 20,000 epoch가 넘어도 비슷한 사진을 생성하였으나, BEGAN의 경우 60,000 epoch가 넘고 나서 실제 사진과 비슷한 사진을 생성하였다. 해당 결과는 [그림 29]와 같다. 해당 사진에서는 혈관, 시신경유두, 황반 모두 관찰할 수 있었으며, 실제 데이터에 존재하는 속눈썹, 관찰자 손가락 등 노이즈도 확인할 수 있었다.



[그림 29] BEGAN 합성 결과 (증상이 있는 데이터 사용)

그러나 모드 붕괴에 빠져 입력이 달라도 똑같은 사진을 생성하는 결과를 보여주었다. BEGAN이 발표된 이후 각 문제를 해결하기 위해 BEGAN v2, v3가 발표되기도 하였다. 그러나 모드 붕괴에 대한 정확한 원인이 존재하지 않으나 BEGAN v2[29] 연구에 따르면 모드 붕괴에 데이터의 수량과 데이터 사이즈가 영향을 미칠 것이라고 한다. 실제 사용된 데이터의 수량을 1/10가량 줄여 사용한 경우 기존의 데이터보다 모드 붕괴에 빨리 빠진 것을 볼 수 있었다. 따라서 해당 실험에서 사용된 데이터의 수량이 적고, 기존 연구들에 사용된 데이터 사이즈보다 실험에 사용된 데이터의 사이즈가 크기 때문에 모드 붕괴가 발생한 것으로 추측할 수 있다.

#### 4.2. 정량적 평가

정량적 평가는 FID, MSE, MAE, PSNR, SSIM을 측정하여 진행하였다. 해당 척도들은 실제 이미지와 생성된 이미지를 비교하여 점수 측정이 가능하므로 선택하였으며, IS의 경우 실제 데이터의 분포를 반영하지 못해 제외하였다. 또한 모델 비교에 사용된 DCGAN, WGAN-GP, BEGAN(0.3), BEGAN(0.5), BEGAN(0.7)으로 나눠 진행했다. 이미지 품질 평가 방법인 MSE, MAE, PSNR, SSIM 점수의 경우 BEGAN에서 실제와 비슷하게 나온 이미지를 추가로 하여 점수를 측정하였다.

각 척도는 실제 데이터와 실제 데이터의 값을 추가로 구하였으며, 해당 점수를 기준으로 모델을 평가하고자 한다.

#### 4.2.1. FID

정량적 평가는 FID를 이용해 평가하였다. 모델별 생성된 데이터와 실제 데이터 세트의 FID를 구하면 다음 <표 5> 같다.

<표 5> 모델별 FID 점수

	실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)
정상	160.8	322.1	323.3	469.1	268.7	237.8
병증	189.1	310.7	307.2	251.3	278.3	252.0

실제 FID 점수를 기준으로 생성기의 능력을 평가하나, Inception 모델의 훈련된 데이터가 ImageNet 데이터로 광각 안저사진과 특징이 다르다. 따라서 실제 데이터와 실제 데이터의 FID를 측정하여, 해당 점수를 기준으로 얼마나 차이가 나는지 비교하려고 한다.

사용된 데이터 세트를 이용해 FID를 구한 경우 정상 데이터 세트가 160.8, 병증 데이터 세트가 189.1 로 나왔다. 점수 차가 가장 적은 모델은 BEGAN(0.7)로 정상 데이터 세트의 경우 77점이 높게 나왔으며, 병증 데이터의 경우 62.3점이 높게 나왔다. 그러나 병증 데이터 세트의 경우 BEGAN(0.7)보다 BEGAN(0.3)의 차이가 62.2로 좀 더 적었다. WGAN-GP의 경우 정상 데이터 세트는 162.5점이 높게 나왔으며, 118.1점이 높게 나왔다. DCGAN의 경우 정상 데이터 세트는 161.3점이 높게 나왔으며, 병증 데이터의 경우 121.6점이 높게 나왔다.

실제 - 실제 데이터의 점수와 직관적인 비교를 위해 실제 - 실제 점수를 기준으로 각 모델의 점수를 백분율로 나타낸 결과가 <표 6>이다.

<표 6> 실제 데이터 기반 FID 백분율

	실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)
정상	100.0%	200.3%	201.1%	291.7%	167.1%	147.9%
병증	100.0%	164.3%	162.5%	132.9%	147.2%	133.3%

두 데이터 세트 모두 50% 이내로 차이 나는 경우는 BEGAN(0.7)밖에 없으며 diversity ratio 값에 따라 DCGAN, WGAN-GP보다 좋은 성능을 보이는 것을 알 수 있다. 또한, 병증 데이터 세트를 사용한 경우 두 모델보다 BEGAN의 세 가지 경우가 차이가 더 적은 것을 알 수 있다.

#### 4.2.2. MSE

MSE는 오차 제곱의 평균으로 해당 값이 작을수록 실제와 비슷한 이미지를 생성한 결과이다. 실제 데이터와 실제 데이터의 점수를 계산할 때 같은 이미지는 제외하고 점수를 계산하였다. 각 이미지별로 점수를 측정하였기 때문에 최댓값과 최솟값, 평균을 제시하였다. 해당 결과는 <표 7>를 참고하면 된다.

MSE 점수를 살펴보면 실제 데이터 세트끼리의 점수도 높은 것을 볼 수 있다. 각 값이 가장 적은 모델은 WGAN-GP이며, BEGAN에서 좋은 결과를 보였던 사진의 최솟값을 제외하면 정상 데이터 세트와 병증 데이터 세트 모두 같은 것을 알 수 있다.

<표 7> 모델별 MSE 점수

		실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)	BEGAN
정상	최대	7,761	10,428	7,778	32,535	12,860	12,236	-
	최소	166	576	475	17,492	2,928	2,439	
	평균	1,542	1,763	1,637	22,709	5,204	4,668	
병증	최대	10,552	12,440	9,842	24,430	18,405	14,261	10,386
	최소	189	525	430	10,449	5,312	1,450	420
	평균	1,480	1,672	1,450	14,409	8,338	3,958	1,607

#### 4.2.3. MAE

MAE는 오차의 절댓값 평균으로 해당 값이 작을수록 실제와 비슷한 이미지를 생성한 것으로 판단할 수 있다. MSE와 같게 실제 데이터 세트의 점수를 계산할 때 같은 이미지는 제외하였다. 각 이미지 별로 점수를 측정하였기 때문에 최댓값과 최솟값, 평균을 제시하였다. 해당 결과는 <표 8>을 참고하면 된다.

MAE 점수도 MSE 점수와 비슷하게 WGAN-GP를 이용하였을 때 점수가 가장 적은 것을 확인할 수 있다. BEGAN의 경우 오토 인코더의 복원 과정을 통해 얻은 데이터의 분포를 이용하기 때문에 생성된 이미지는 실제와 유사해 보이나, 각 픽셀을 분석한 경우에는 WGAN-GP보다 차이가 크게 보이는 것으로 추측할 수 있다.

<표 8> 모델별 MAE 점수

		실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)	BEGAN
정상	최대	90.0	141.5	83.4	251.7	150.7	147.0	-
	최소	12.5	36.3	23.7	181.6	71.9	64.8	
	평균	34.9	58.7	40.4	211.0	97.8	92.7	
병증	최대	99.0	163.9	95.6	213.5	183.3	157.4	112.3
	최소	14.0	33.0	22.3	138.0	98.4	49.2	32.3
	평균	32.9	56.5	36.3	163.0	122.9	83.9	47.4

#### 4.2.4. PSNR

PSNR은 최대 신호 대 잡음비로 이미지를 압축하거나 재구성하였을 때 품질을 비교하는 척도이다. 해당 척도를 GAN에 적용하면 PSNR 값이 클수록 이미지의 품질이 좋다. 해당 결과는 <표 9>을 참고하면 된다.

PSNR 점수를 비교하면 BEGAN에서 좋은 결과를 보였던 경우의 최댓값을 제외하고 WGAN-GP에서 값이 제일 큰 것을 알 수 있다. WGAN-GP와 BEGAN의 좋은 결과를 비교하면 값의 차이가 크지 않다.

BEGAN의 경우 diversity ratio의 값이 증가할수록 점수가 높아지는 것을 알 수 있다. BEGAN의 경우 diversity ratio의 값이 증가할수록 합성되는 이미지의 다양성이 증가하고 품질이 떨어진다. 그러나 실제 데이터 세트에서 diversity ratio 값이 증가하여도 점수가 큰 것을 알 수 있다.

<표 9> 모델별 PSNR 점수

		실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)	BEGAN
정상	최대	25.9	20.5	21.3	5.7	13.4	14.2	-
	최소	9.2	7.9	9.2	3.0	7.0	7.2	
	평균	17.4	16.2	16.6	4.5	11.0	11.5	
병증	최대	25.3	20.9	21.7	7.9	10.8	16.5	21.8
	최소	7.8	7.1	8.1	4.2	5.4	6.5	7.9
	평균	17.7	16.6	17.1	6.5	8.9	12.4	16.7

#### 4.2.5. SSIM

SSIM은 PSNR이 이미지의 구조적 정보를 반영하지 못하기 때문에 만들어진 평가 척도이다. SSIM의 경우 이미지가 같은 경우 1 값이 나오고 값이 클수록 실제 이미지와 구조적으로 유사하다는 것을 알 수 있다. 해당 결과는 <표 10>을 참고하면 된다.

SSIM의 경우 BEGAN에서 좋은 결과를 생성했던 사진의 최댓값이 0.95로 훈련 데이터와 유사하나 같지 않은 것을 알 수 있다. 또한, 위의 평가 척도들과 유사하게 WGAN-GP의 결과가 큰 값이 나온 것을 알 수 있다. 그리고 BEGAN의 경우 diversity ratio 값에 따라 값이 증가하는 것을 알 수 있으며, 값의 증가 폭은 0.5일 때와 0.7일 때 크지 않은 것을 알 수 있다.

<표 10> 모델별 SSIM 점수

		실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)	BEGAN
정상	최대	0.90	0.47	0.77	0.41	0.65	0.66	-
	최소	0.65	0.26	0.53	0.25	0.38	0.37	
	평균	0.80	0.40	0.70	0.33	0.54	0.54	
병증	최대	0.90	0.52	0.78	0.60	0.60	0.66	0.95
	최소	0.54	0.18	0.52	0.29	0.27	0.29	0.45
	평균	0.78	0.40	0.68	0.50	0.51	0.54	0.78

정량적 평가 결과를 정리하면 다음과 같다. FID의 경우 BEGAN이 가장 성능이 좋게 나왔으나 이미지 품질을 평가한 경우 WGAN-GP가 가장 실제와 비슷한 것으로 나타났다. BEGAN과 WGAN-GP 모두 Wasserstain distance를 이용하여 목적 함수를 설정하나, BEGAN의 경우 오토인코더의 결과를 이용하며, WGAN-GP는 데이터의 확률 분포를 이용한다는 차이가 존재한다. BEGAN의 경우 오토인코더의 결과를 이용하기 때문에 실제 이미지의 특징을 학습할 수 있었던 것으로 보이며, WGAN-GP의 경우 실제 데이터 분포와 비슷해지려고 하므로 픽셀별 계산을 진행하는 이미지 품질 평가에서 좋은 결과를 얻은 것으로 보인다.

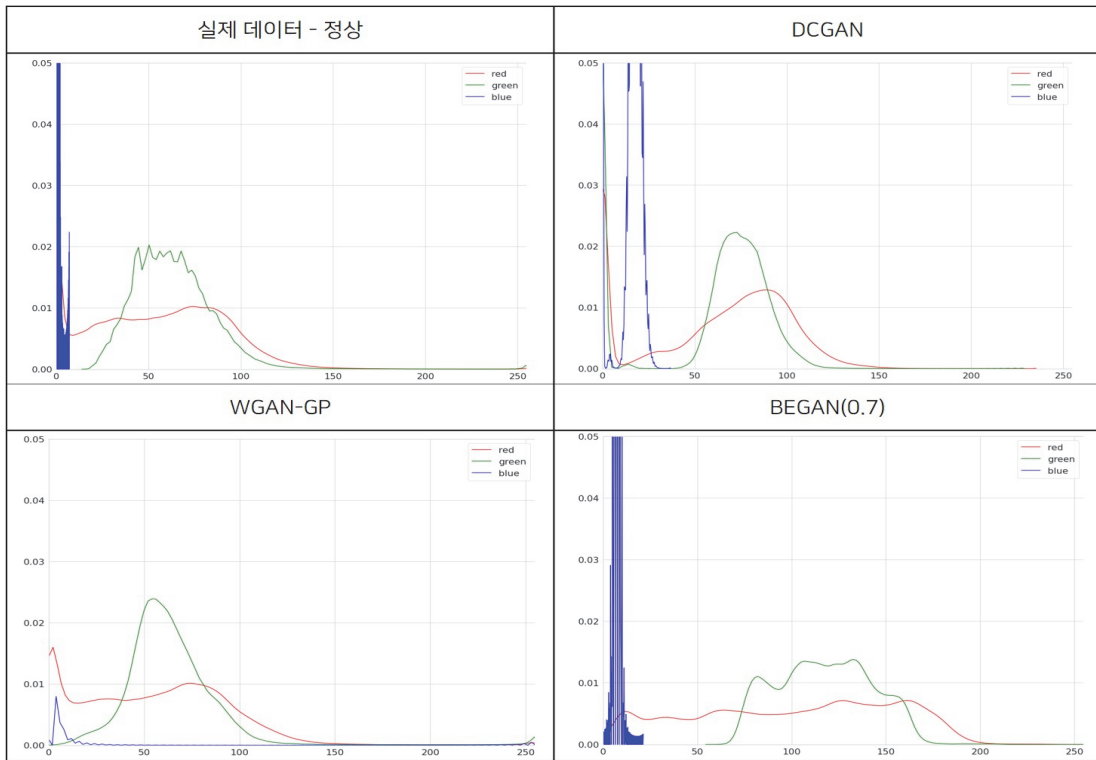
## V. 결론

본 논문에서는 다양한 GAN 모델을 이용하여 광각 안저사진 합성 연구를 진행하였으며, 다양한 평가 척도를 이용하여 결과를 분석·비교하였다. 합성 결과 분석은 커널 밀도 추정 방법을 이용하여 데이터 분포를 분석하였다. 결과 비교를 진행하기 위해 정량적 평가와 정성적 평가를 했으며, 각 평가 결과를 종합하였을 때 BEGAN이 좋은 결과를 얻었다. 합성 결과에 따른 커널 밀도 추정 결과, 평가 척도 결과를 정리하면 다음과 같다.

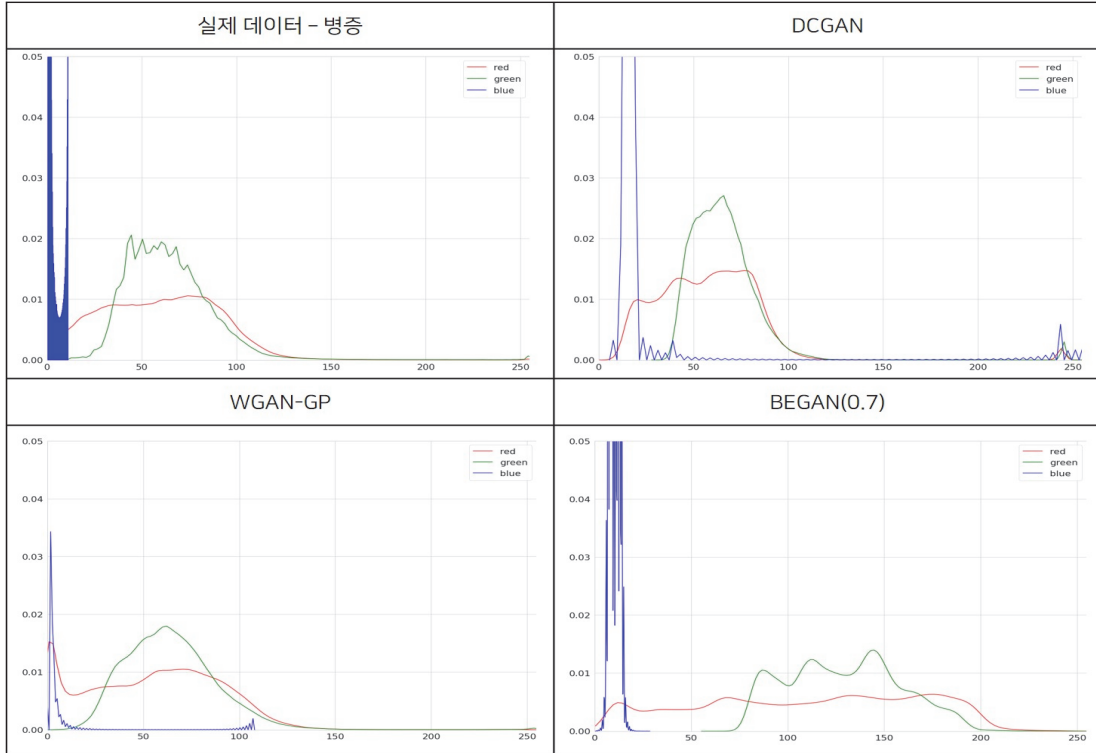
DCGAN, WGAN-GP, BEGAN을 이용하여 광각 안저사진을 합성한 결과를 이용하여 커널 밀도 추정을 하였으며, 합성 결과의 R, G, B 데이터 분포가 실제 데이터의 분포와 어떻게 다른지 비교하였다. 훈련 말기의 결과와 실제 데이터의 분포는 [그림 30]과 [그림 31]을 통해 다시 한 번 확인할 수 있다. 색상별로 초록색의 경우 DCGAN과 WGAN-GP가 데이터 분포가 유사하였으나 DCGAN이 각 값의 밀도 정도가 실제보다 더 높고, WGAN-GP는 데이터 분포가 좀 더 넓은 범위로 퍼져있다. 빨간색의 경우 WGAN-GP가 실제와 가장 유사하고, 파란색의 경우 BEGAN이 가장 유사하였다.

각 모델의 목적 함수가 다르므로 합성된 데이터 분포의 양상도 다르다는 것을 알 수 있었으며, 데이터에 영향을 미치는 데이터의 색도 다르다는 것을 알 수 있었다. BEGAN에서 판별기와 생성기의 훈련 균형을 맞추기 위해 diversity ratio를 이용하는 것처럼 GAN에서 이미지 분포를 학습할 때 각 색의 중요도를 하이퍼 파라미터로 추가한다면 실제와 좀 더 유사한 데이터 분포를 볼 수 있을 것으로 기대된다.





[그림 30] 정상 데이터를 이용한 합성 결과 데이터 분포



[그림 31] 병증 데이터를 이용한 합성 결과 데이터 분포

세 모델을 비교하기 위해 정성적, 정량적 평가를 진행하였으며, 정량적 평가의 경우 FID와 이미지 품질 평가 척도(MSE, MAE, PSNR, SSIM)을 이용하였다. 정성적 평가와 FID의 결과 BEGAN이 가장 좋았으며, 이미지 품질 평가 척도에서는 WGAN-GP가 가장 좋은 결과를 보여주었다.

정성적 평가는 안저사진에서 관찰되어야 하는 요소인 전체적인 윤곽, 혈관, 시신 경유두, 황반, 증상의 유무를 통해 판단하였다. BEGAN을 이용한 경우 100,000 epoch가 넘었을 때 실제와 유사한 이미지가 생성되었으나 모드 붕괴에 빠져 같은 이미지만을 생성하였다. 같은 이미지를 생성하였지만 BEGAN을 이용하였을 때 가장 좋은 결과를 보여주었다.

세 모델의 정량적 평가를 정리하면 정상 데이터를 사용한 경우 <표 11>이며, 병증 데이터를 사용한 경우 <표 12>에 해당한다. <표 11>과 <표 12>에서는 FID 점수를 제외한 평가 척도에서는 평균값을 표시하였다. FID 점수의 경우 ImageNet 데이터를 이용하여 미리 훈련된 Inception 모델을 이용하여 데이터 분포를 구하고, 실제 데이터 분포와 거리를 측정한다. FID를 이용한 성능 비교를 위해서는 비교군이 존재해야 그러나 본 연구와 유사한 연구가 없었기 때문에 실제 데이터 세트와 실제 데이터 세트의 FID 점수를 기준으로 평가하였다. FID 점수는 데이터 분포의 거리 차를 구하기 때문에 값이 작을수록 실제와 유사하다는 것을 나타낸다.

FID 점수 차가 가장 적은 모델은 BEGAN이었다. 또한, BEGAN의 diversity ratio 값에 따라 각 모델의 FID 점수가 다르게 나오는 것을 확인할 수 있었다. 정상 데이터의 경우 diversity ratio의 값이 커질수록 FID 점수가 작아지는 것을 볼 수 있었으나, 병증 데이터의 경우 diversity ratio 값이 0.3과 0.7인 경우 FID 점수가 작은 값으로 측정되었다.

이미지 품질 평가의 경우 MSE, MAE, PSNR, SSIM을 이용하였다. MSE와 MAE는 픽셀의 오차를 이용하기 때문에 값이 작을수록 좋은 결과를 나타내며, PSNR의 경우 값이 클수록 실제와 비슷하다는 것을 나타낸다. SSIM은 0 ~ 1 사이의 값을 가지며, 같은 이미지를 이용하여 SSIM 점수를 구하였을 때, 1의 값이 도출된다.

MSE, MAE, PSNR, SSIM 점수를 살펴보면 WGAN-GP 모델이 가장 좋은 결과를 보여주었다. 이미지 품질 평가는 픽셀의 오차를 이용하여 측정하기 때문에 Wasserstein distance를 이용하는 WGAN-GP가 좋은 결과를 보여준 것으로 분석

된다. BEGAN도 Wasserstein distance를 이용하지만 실재가 아닌 오토인코더 손실을 이용하여 구하기 때문에 실제 픽셀값은 WGAN-GP와 비교하였을 때 차이가 존재하는 것으로 추측할 수 있다. 병증 데이터를 이용해 이미지를 합성하였을 때 BEGAN 모델 중 실재와 비슷한 이미지가 도출되었다. 해당 이미지를 이용해 이미지 품질 평가를 진행하였을 때, WGAN-GP와 DCGAN의 중간 범위에 해당하는 결과를 보여주었다.

<표 11> 모델별 정량적 평가 결과 (정상 데이터)

		실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)	BEGAN
정상	FID	160.8	322.1	323.3	469.1	268.7	237.8	-
	MSE	1,542	1,763	1,637	22,709	5,204	4,668	
	MAE	34.9	58.7	40.4	211.0	97.8	92.7	
	PSNR	17.4	16.2	16.6	4.5	11.0	11.5	
	SSIM	0.80	0.40	0.70	0.33	0.54	0.54	

<표 12> 모델별 정량적 평가 결과 (병증 데이터)

		실제-실제	DCGAN	WGAN-GP	BEGAN(0.3)	BEGAN(0.5)	BEGAN(0.7)	BEGAN
병증	FID	189.1	310.7	307.2	251.3	278.3	252.0	-
	MSE	1,480	1,672	1,450	14,409	8,338	3,958	1,607
	MAE	32.9	56.5	36.3	163.0	122.9	83.9	47.4
	PSNR	17.7	16.6	17.1	6.5	8.9	12.4	16.7
	SSIM	0.78	0.40	0.68	0.50	0.51	0.54	0.78

정량적 평가 척도 결과도 중요하지만, 실제 딥러닝 연구에 사용되기 위해서는 정성적인 평가 또한 중요하다. 따라서 GAN을 이용해 광각 안저사진을 합성하는 경우 세 모델 중에서 BEGAN이 가장 적합하며, BEGAN의 diversity ratio 값이 0.7인 경우 가장 좋은 결과를 보여주었다.

GAN의 특성상 데이터 세트에 따른 최적의 하이퍼 파라미터를 설정하는 것이 중요하다. 따라서 향후 연구에서는 BEGAN 모델을 이용해 하이퍼 파라미터를 최적화한다면 실제 광각 안저사진에 가까운 품질의 사진을 합성할 수 있을 것이라 기대

한다. 또한, 데이터 분포를 보았을 때 BEGAN의 경우 빨간색 데이터 분포가 실제와 유사하지 않았기 때문에 이러한 부분을 보완할 방법을 추가한다면 좀 더 유사한 이미지를 만드는 데 도움이 될 것이라 기대한다. 유사한 이미지를 만든 후 평가할 수 있는 평가 척도 또한 필요하다. 본 논문에서는 다양한 평가 척도를 이용하여 합성된 이미지를 평가하였다. 그러나 일정 기준 이상이 됐을 때 딥러닝 연구에 사용할 수 있는 사진이라는 것은 판단하지 못하였다. 따라서 합성된 이미지를 딥러닝 연구에 사용하기 위해서는 일정한 기준과 그에 맞는 평가 척도가 필요하다.

[참고문헌]

- [1] Falavarjani, K. G., Tsui, I., & Sadda, S. R. (2017). Ultra-wide-field imaging in diabetic retinopathy. *Vision research*, 139, 187-190.
- [2] 공안과. "특집: 당뇨병과 안과질환 -만화로 쉽게 이해하는 당뇨망막병증" 당뇨병 VOL.255 NO.- (2011):12-17
- [3] Price, L. D., Au, S., & Chong, N. V. (2015). Optomap ultrawide field imaging identifies additional retinal abnormalities in patients with diabetic retinopathy. *Clinical ophthalmology (Auckland, NZ)*, 9, 527.
- [4] 김경운. (2014). Optomap 장비를 이용한 광각안저검사의 진단적 효용성. *임상생리검사학회 초록집*, 2014(2), 23-26.
- [5] 이현정. (2017. 02. 01). 3대 실명(失明)질환 젊다고 안심하면 안 된다. *헬스조선*.
- [6] Nagasawa, T., Tabuchi, H., Masumoto, H., Enno, H., Niki, M., Ohara, Z., ... & Mitamura, Y. (2019). Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy. *International ophthalmology*, 39(10), 2153-2159.
- [7] Song KD, Kim M, Do S. (2019). The Latest Trends in the Use of Deep Learning in Radiology Illustrated Through the Stages of Deep Learning Algorithm Development. *J Korean Soc Radiol*, 80(2), 202-212.
- [8] Nagasawa, T., Tabuchi, H., Masumoto, H., Enno, H., Niki, M., Ohsugi, H., & Mitamura, Y. (2018). Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes. *PeerJ*, 6, e5696.
- [9] Ohsugi, H., Tabuchi, H., Enno, H., & Ishitobi, N. (2017). Accuracy of

- deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. *Scientific reports*, 7(1), 1-4.
- [10] Masumoto, H., Tabuchi, H., Nakakura, S., Ohsugi, H., Enno, H., Ishitobi, N., ... & Mitamura, Y. (2019). Accuracy of a deep convolutional neural network in detection of retinitis pigmentosa on ultrawide-field images. *PeerJ*, 7, e6900.
- [11] Nagasato, D., Tabuchi, H., Ohsugi, H., Masumoto, H., Enno, H., Ishitobi, N., ... & Mitamura, Y. (2019). Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion. *International journal of ophthalmology*, 12(1), 94.
- [12] Li, Z., Guo, C., Nie, D., Lin, D., Zhu, Y., Chen, C., ... & Yang, Y. (2020). Development and evaluation of a deep learning system for screening retinal hemorrhage based on ultra-widfield fundus images. *Translational Vision Science & Technology*, 9(2), 3-3.
- [13] Li, Z., Guo, C., Nie, D., Lin, D., Zhu, Y., Chen, C., ... & Xiao, H. (2019). A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widfield fundus images. *Annals of Translational Medicine*, 7(22).
- [14] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501-509, 2004.
- [15] Hoover, A. D., Kouznetsova, V., & Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3), 203-210.
- [16] Costa, P., Galdran, A., Meyer, M. I., Niemeijer, M., Abràmoff, M., Mendonça, A. M., & Campilho, A. (2017). End-to-end adversarial retinal image synthesis. *IEEE transactions on medical imaging*,

37(3), 781-791.

- [17] Guibas, J. T., Viridi, T. S., & Li, P. S. (2017). Synthetic medical images from dual generative adversarial networks. arXiv preprint arXiv:1709.01872.
- [18] Zhao, H., Li, H., Maurer-Stroh, S., & Cheng, L. (2018). Synthesizing retinal and neuronal images with generative adversarial nets. *Medical image analysis*, 49, 14-26.
- [19] Iqbal, T., & Ali, H. (2018). Generative adversarial network for medical images (MI-GAN). *Journal of medical systems*, 42(11), 231.
- [20] Biswas, S., Rohdin, J., & Drahanický, M. (2019, July). Synthetic Retinal Images from Unconditional GANs. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2736-2739). IEEE.
- [21] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).
- [22] Lucic, M., Kurach, K., Michalski, M., Gelly, S., & Bousquet, O. (2018). Are gans created equal? a large-scale study. In *Advances in neural information processing systems* (pp. 700-709).
- [23] Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical image analysis*, 101552.
- [24] 김한웅. (2018). Generative adversarial network을 이용한 Alzheimer's disease의 분류. 연세대학교 대학원 생체공학협동과정 전기전자공학전공.
- [25] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
- [26] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan.

arXiv preprint arXiv:1701.07875.

- [27] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. In Advances in neural information processing systems (pp. 5767-5777).
- [28] Berthelot, D., Schumm, T., & Metz, L. (2017). Began: Boundary equilibrium generative adversarial networks. arXiv preprint arXiv:1703.10717.
- [29] Chang, C. C., Hubert Lin, C., Lee, C. R., Juan, D. C., Wei, W., & Chen, H. T. (2018). Escaping from collapsing modes in a constrained space. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 204-219).



<ABSTRACT>

## Comparative Study of Generative Adversarial Networks Using Ultra-wide-field Fundus Image

ARa Ko

Graduate School of Jeju National University  
(Major in Convergence Education Software)

Supervised by Professor Jungwon Cho

In actual medical settings recently, due to an advancement of medical imaging technologies, the conventional fundus cameras are being replaced with the ultra-wide field retinal imaging devices that enable to take images easier and provide the broadened imaging range. However, due to difficulties in data collection process, most of the deep learning studies related to ophthalmologic diseases are still being conducted based on the fundus images by the conventional fundus cameras.

The performance of deep learning is associated with data volume; accordingly, the data augmentation is used to amplify the data for performance enhancement in case of data shortage. Data augmentation includes the general data augmentation method that applies rotation or reversal to images and the method that synthesizes new images using GAN.

It is a big challenge to evaluate the performance of each model because there are various models in GAN, in other words, because the

dataset used in each model is different and no suitable evaluation criteria are available. In addition, the data used for GAN model research cannot reflect the resolution and complexity metric of data that are used in actual synthesis. In this context, this study synthesizes the ultra-wide field retinal images taken by an ultra-wide field retinal imaging device using various GAN models and comparatively analyzes the synthesis results.

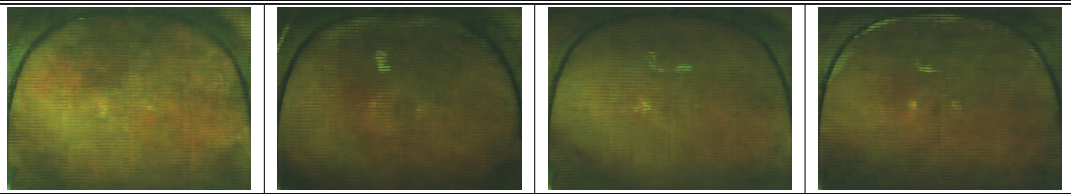
The data for ultra-wide field retinal images taken by an ultra-wide field retinal imaging device, which are limited only with the diabetic retinopathy, were collected and labelled through a collaboration with the ophthalmology department of Jeju National University Hospital. The GAN model used in this study has utilized DCGAN, WGAN-GP and BEGAN that are all frequently used in the studies for medical image synthesis. Some hyper-parameters in each model were set to be identical and other parameter values were set to the criteria of the studies that each model was reported.

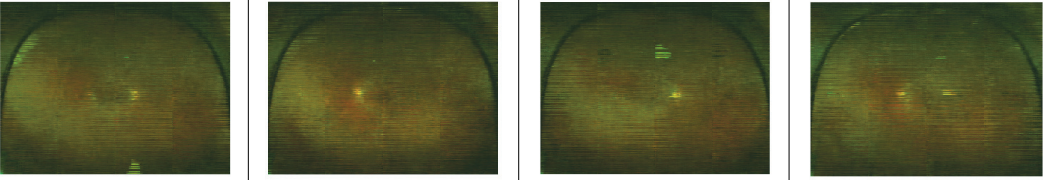
This study conducted the qualitative and quantitative evaluations on the synthesized ultra-wide field retinal images. The qualitative evaluation was proceeded for the availability in observation of optic disc or macular etc. Both FID from those images and image quality evaluation scale were used for the quantitative evaluation. The FID obtained the best evaluation result when BEGAN model was used and WGAN-GP model exhibited a good result in the image evaluation. When comparing the results after proceeding the qualitative and quantitative evaluations, it was thought that BEGAN was most appropriate. Moreover, the similar results to actual ultra-wide field retinal images were obtained when BEGAN repeated the training over and over again. It is expected to obtain even more sophisticated results if the hyper-parameters of BEGAN are adjusted according to the characteristics of each dataset.

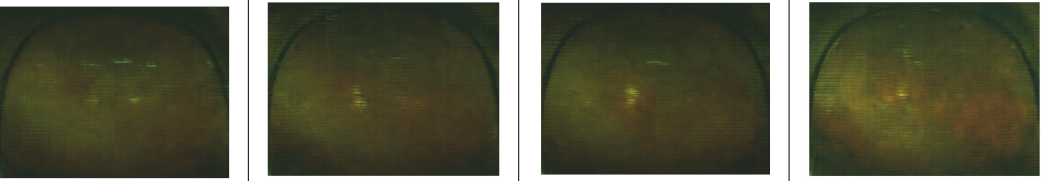
[부록]

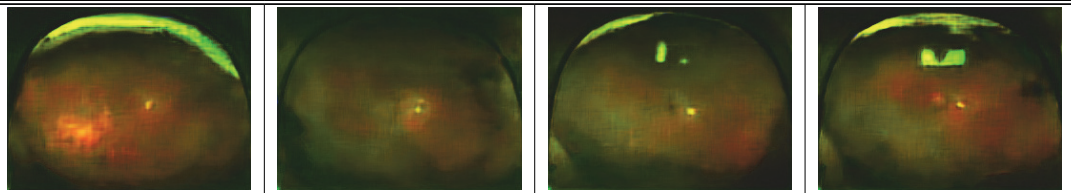
본 연구에서는 모델 비교를 위해 하이퍼 파라미터를 유사하게 설정하여 이미지를 합성하였다. 그 외에 하이퍼 파라미터를 변경하여 사용한 경우를 부록으로 제시한다. 세 모델의 구조는 같게 이용하였으며, 각 경우에 합성된 결과를 제시한다.

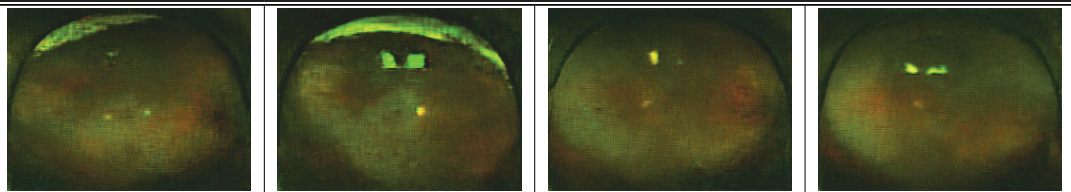
DCGAN의 경우 GAN, 생성기(Generator), 판별기(Discriminator)의 최적화 함수가 다르므로 각각 제시하였다. WGAN-GP와 BEGAN의 경우 생성기와 판별기의 최적화 함수가 같다. BEGAN은 Diversity ratio 값을 Gamma로 표기하였으며, 목적 함수를 계산하는 데 있어 필요한 k 값을 설정하기 위한 학습률(learning rate)을 learning rate for k로 표기하여 사용된 값을 제시하였다.

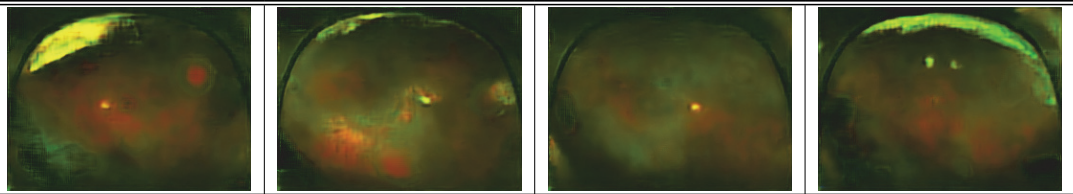
모델	DCGAN	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	20,000	이미지 사이즈	512 * 384
배치 사이즈	16	잠재공간 사이즈	100
GAN 최적화	GAN : RMSprop - learning rate = 0.0005, decay = 3e-7 Generator : Discriminator :		
<b>합성 결과</b>			
			

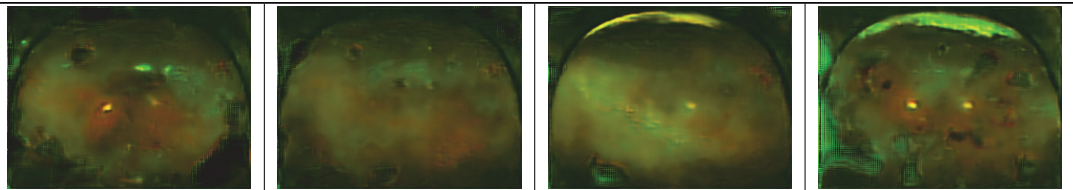
모델	DCGAN	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	20,000	이미지 사이즈	400 * 300
배치 사이즈	16	잠재공간 사이즈	50
GAN 최적화	GAN : RMSprop - learning rate = 0.0005, decay = 3e-7 Generator : RMSprop - learning rate = 0.0001, decay = 3e-7 Discriminator : RMSprop - learning rate = 0.0001, decay = 3e-7		
<b>합성 결과</b>			
			

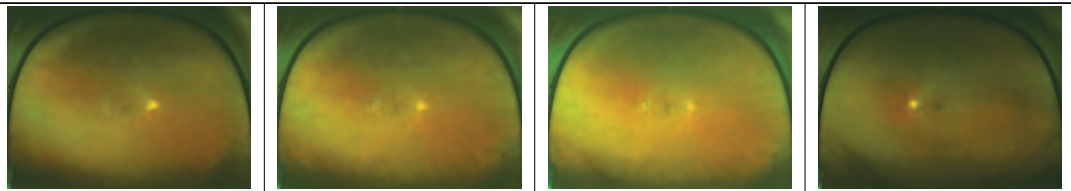
모델	DCGAN	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	20,000	이미지 사이즈	512 * 384
배치 사이즈	16	잠재공간 사이즈	100
GAN 최적화	GAN : RMSprop - learning rate = 0.0005, decay = 3e-7 Generator : RMSprop - learning rate = 0.0001, decay = 3e-7 Discriminator : RMSprop - learning rate = 0.0001, decay = 3e-7		
<b>합성 결과</b>			
			

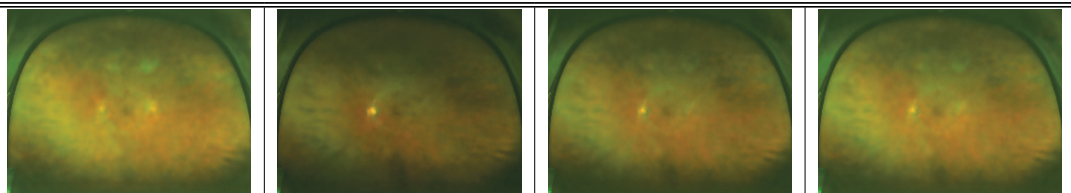
모델	WGAN-GP	데이터	정상
<b>하이퍼 파라미터</b>			
반복 횟수	100,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	100
GAN 최적화	Adam - learning rate = 0.0001, beta_1 = 0, beta_2 = 0.9		
<b>합성 결과</b>			
			

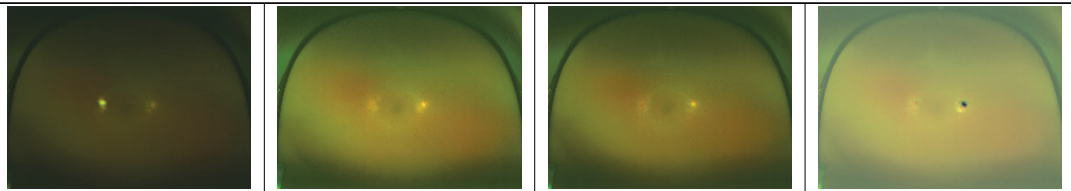
모델	WGAN-GP	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	100,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	100
GAN 최적화	RMSprop - learning rate = 0.00005		
<b>합성 결과</b>			
			

모델	WGAN-GP	데이터	정상
<b>하이퍼 파라미터</b>			
반복 횟수	100,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	100
GAN 최적화	Adam - learning rate = 0.0001, beta_1 = 0.5, beta_2 = 0.9		
<b>합성 결과</b>			
			

모델	WGAN-GP	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	100,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	100
GAN 최적화	Adam - learning rate = 0.0001, beta_1 = 0.5, beta_2 = 0.9		
<b>합성 결과</b>			
			

모델	BEGAN	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	100,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	32
GAN 최적화	gamma : 0.5, learning rate for k = 0.001 Adam - learning rate = 0.0001		
<b>합성 결과</b>			
			

모델	BEGAN	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	400,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	100
GAN 최적화	gamma : 0.5, learning rate for k = 0.001 Adam - learning rate = 0.0001		
<b>합성 결과</b>			
			

모델	BEGAN	데이터	정상
<b>하이퍼 파라미터</b>			
반복 횟수	100,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	100
GAN 최적화	gamma : 0.7, learning rate for k = 0.001 Adam - learning rate = 0.0001		
<b>합성 결과</b>			
			

모델	BEGAN	데이터	병증
<b>하이퍼 파라미터</b>			
반복 횟수	800,000	이미지 사이즈	256 * 192
배치 사이즈	8	잠재공간 사이즈	32
GAN 최적화	gamma : 0.7, learning rate for k = 0.001 Adam - learning rate = 0.0001		
<b>합성 결과</b>			
