**Page 1**

* 안녕하십니까, 전자전기컴퓨터공학과 재학중인 김현규입니다.
* 지금부터 심층 신경망을 이용한 근적외선 채색 방법에 관한 연구 심사 발표 시작하겠습니다.

**Page 2**

* 목차는 주제 소개, 제안 방법, 실험 결과, 결론 순으로 작성했으며, 순서대로 설명 드리겠습니다.

**Page 3**

* 근적외선 영상은 700 – 1200 나노미터 대역의 빛을 영상화 한 것으로, 일반 영상보다 더 멀리 물체와 디테일을 포착하는 능력을 갖습니다.
* 왼쪽 그림에서, 아래 적외선 카메라를 사용하여 더 먼 물체의 빛을 반사하여 자세한 영역을 촬영하는 것을 알 수 있습니다.
* 오른쪽 그림은 예시로, 위는 근적외선 영상, 아래는 일반 영상이며, 디테일을 더욱 포착하는 것을 알 수 있습니다.
* 현재, 근적외선 영상은 영상 감시, 안개 낀 영상 촬영 등 다양한 분야에서 사용되고 있습니다.

**Page 4**

* 관련 연구로 기존 흑백 영상-+ 채색기법이 있습니다. 흑백 영상 채색 기법은 왼쪽 그림을 보시면 다양한 흑백 영상으로부터 컬러 영상을 복원하는 기술이고, 최근 딥러닝을 이용해 연구되고 있습니다.
* 근적외선 영상 채색 기법은 이 주제의 특정한 영역으로 많은 공통점을 갖습니다.
* 오른쪽 위 그림은 흑백 영상 채색기법 초창기 모델로, CNN 구조만 사용했습니다.
* 아래 그림은 근적외선 영상 채색기법 초창기 모델로, CNN과 전처리, 후처리로 구성되있으며, 특히 후처리에 디테일을 추가한 방법이 눈에 띕니다. 저는 다른 방식으로 디테일을 개선하는 방법을 연구하였습니다.

**Page 5**

* 제안 방법 설명 드리기 전, 먼저 baseline을 소개하겠습니다.
* Baseline은 2020 ECCV에서 소개된 근적외선 영상 채색 기법에 대한 모델입니다.
* 그림과 같이, 모델 구조는 1개의 인코더와 2개의 병렬 디코더로 구성됩니다.
* ResNet 구조를 사용하며, U-Net에서 사용된 skip connection을 적용했습니다.
* 또한, texture와 색 정보간 상관관계를 위한 State Synchronization module이 이용됩니다.
* 왼쪽 위 그림을 보시면, 이는 CIE Lab format을 사용하고, A가 L, B가 ab이며, 코사인 유사도와 가우시안 커널을 통과해 similairy map을 생성하게 되고, 기존의 L과 ab에 곱해주게 됩니다. 이를 디코더에 사용해 영상 복원 성능을 높였습니다.

**Page 6**

* 다음은 VAE와 VQ-VAE를 적용한 제안 방법 설명 드리겠습니다. VAE는 비지도 학습 방식을 사용하여, 인코딩한 정보를 확률 분포 형태로 학습하는 모델입니다.
* 또한, VQ-VAE는 인코딩한 정보를 이산 값으로 변환하고 codebook에 매핑하는 모델로 VAE보다 좋은 성능을 가진 모델입니다.
* 이러한 VAE와 VQ-VAE는 입력의 중요 정보만을 추출하여 새로운 데이터를 생성하는 generative model의 특성을 가지고 있고, 이것을 baseline에 적용하여 성능을 높이고자 하였습니다.
* 왼쪽 그림을 보시면, VAE의 아이디어를 이용해 baseline의 인코딩 정보를 확률 분포 형태로 학습하게 됩니다.
* 오른쪽 그림을 보시면, baseline은 그대로 유지하되, VQ-VAE 모델을 추가하여 인코딩한 정보를 baseline의 인코더 끝에 concatenation 했습니다.
* 2가지 방법 모두 VAE 혹은 VQ-VAE의 특성을 baseline과 융합해 성능을 향상시키는 것을 목표로 적용했습니다.
* 또한, baseline과 다르게 3개의 병렬 디코더를 사용하였으며, baseline의 State synchronization module과 조금 다른 module을 설계했습니다. 기존의 L과 ab 대신 L과 a, L과 b 각각을 사용하면 색차 정보가 더 풍부해질 것이라 생각했습니다.

**Page 7**

* 마지막 제안 방법은 방향을 바꿔 VAE를 사용하지 않은 모델입니다.
* Baseline 모델은 detail이 사라지고 색의 일관성이 떨어지는 문제가 발생하게 되어, 이를 해결하고자 canny filter의 edge detection을 활용하고 feature fusion을 통해 영상을 융합하여 성능을 개선하고자 했습니다.
* 전체 구조는 왼쪽 그림처럼 3개의 동일한 baseline 구조, edge detection, feature fusion module로 구성됩니다.
* Feature fusion module은 2021 WACV에 소개된 MS-CAM을 기반으로 했습니다. MS-CAM은 Multi-scale channel attention module의 약자로 서로 다른 object를 효과적으로 포착하기 위해 설계되었습니다. 그림의 오른쪽 경로는 global attention과 local attention으로 이루어지고 이를 통해 전역 정보와 지역적 정보 모 획득할 수 있습니다. 그 후, 두 정보는 합쳐지게 되고, sigmoid를 지나 입력에 곱해지게 됩니다.
* 그러나, MS-CAM은 모든 채널을 같이 사용하며, 이 때 정보들이 모두 뒤섞이는 단점이 있습니다. 따라서 제안하는 feature fusion module은 MS-CAM을 2번 사용했습니다. 오른쪽 그림처럼 앞서 2개의 baseline으로 출력한 2개의 컬러 영상에서 L과 ab를 분리하여 사용합니다. 즉, 서로의 L 성분과 ab 성분을 따로 MS-CAM에 사용해 feature fusion을 수행합니다.

**Page 8**

* 목적 함수는 다음 3가지 loss를 합해서 사용했습니다.
* 첫번째로, DSSIM loss입니다. SSIM을 이용하고, 영상의 구조적 차이를 위해 사용합니다.
* 두번째로, L1 loss이고, pixel 값 추정을 위해 사용합니다.
* 마지막으로, perceptual loss입니다. pre-trained VGG의 출력간 차이를 줄이는 형태이고, 시인성, 즉 얼마나 실제 영상처럼 보이는지를 개선합니다.
* 각 모델별 사용한 loss를 아래 표로 정리했습니다.
* CIE Lab format에서 L은 밝기와 디테일 정보를, ab는 색차 정보를 나타냅니다.
* 따라서, baseline을 보시면 L에는 DSSIM loss를 사용하고, ab에는 L1 loss를 적용합니다. 2가지 loss 각각 L과 ab에만 따로 적용됩니다. VAE를 적용한 모델은 L1 loss만 적용했으며, VQ-VAE 를 적용한 모델은 a와 b에 따로 L1 loss를 적용했습니다. Edge와 feature fusion을 적용한 모델은 baseline과 동일하나, (전 모델 그림으로 돌아가기) 모델의 출력 3가지에 순서대로 적용됩니다. 기존 baseline을 edge로부터의 출력, feature fusion module, 그리고 baseline 하나를 더 추가하여 강화하는 형태로 사용했습니다. 마지막으로 람다는 모두 10으로 설정했습니다.

**Page 9**

* 다음은 실험 결과에서 먼저 데이터셋부터 설명 드리겠습니다.
* 데이터세트는 Video surveillance in the day의 약자인 VSIAD 데이터셋을 사용했습니다. 왼쪽 그림을 보시면 VSIAD는 주간 모드의 VNIR2VC와 야간 모드의 NIR2VC 2가지 이미지 pair로 구성됩니다. VNIR2VC는 컬러 영상과 근적외선 영상이 합쳐진 데이터이며, NIR2VC는 근적외선 영상만 포함됩니다.
* 또한, 오른쪽 그림과 같이 VSIAD 데이터셋은 Position 0, 1, 2, 3의 4개 지점에 카메라를 설치하고 촬영했습니다.
* 총 20,000 장으로 구성되며, train, validation, test를 3:1:1 비율로 나눠 사용했습니다.

**Page 10**

* 실험 결과는 다음과 같이 표로 나타냈습니다.
* 비교 모델로 기존 방법인 pix2pix, pix2pixHD 그리고 State-of-art 모델 baseline 3가지를 사용했으며, 평가 지표로는 PSNR, SSIM, LPIPS를 사용했습니다.
* 먼저 표를 보시면 VNIR2VC에 포함된 컬러 정보로 인해 기본적으로 VNIR2VC의 결과가 NIR2VC보다 훨씬 좋은 것을 알 수 있습니다.
* pix2pix와 pix2pixHD는 GAN 알고리즘이 적용된 모델인데, pix2pix는 초창기 모델로 매우 낮은 성능을 보여주었습니다. 그러나, 개선된 모델인 pix2pixHD는 baseline에 매우 근접한 성능을 보였으며, NIR2VC의 PSNR은 오히려 baseline보다 높게 측정되었습니다.
* SSIM에서는 baseline이 가장 높게 측정되며, 이는 DSSIM loss가 적용되어 출력 영상의 구조적 특징이 타겟 영상과 매우 유사함을 나타냅니다.
* LPIPS는 출력 영상이 실제 영상처럼 자연스러운지를 수치로 나타낸 평가 지표이고, 낮을수록 좋은 성능을 나타냅니다. LPIPS에서는 오히려 pix2pix가 가장 좋은 성능을 보였습니다.
* 마지막으로, edge와 feature fusion을 적용한 모델은 NIR2VC의 PSNR에서 baseline보다 높게 측정됩니다.
* 요약하면, VAE와 VQ-VAE를 적용한 모델은 당초 기대와 예상보다는 좋은 성능 구현이 어려웠으며, edge와 feature fusion을 적용한 모델은 baseline보다 PSNR이 약 0.8 상승하였습니다.

**Page 11**

* 다음은 출력 영상들을 비교해 보여드리겠습니다.
* 먼저 VNIR2VC입니다. 출력 영상은 차례대로 입력, 비교 모델, 3가지의 제안 방법들과 타겟입니다.
* Pix2pix는 약간의 영상 왜곡이 발생하고, pix2pixHD는 가장 선명한 영상을 생성합니다. Baseline은 조금 어두운 영상을 생성합니다.
* VAE 적용 방법은 상단에 line artifact를 생성됩니다. 이는 overfitting 결과로서, 모든 출력에서 생성되었습니다. 또한, 신호등 색이 다르게 생성됩니다.
* VQ-VAE 적용 방법은 신호등 색이 파랗게 생성되는 오류가 나타납니다.
* 마지막 제안 방법은 조금 밝은 영상을 생성하고, 오른쪽에 색이 번지는 현상이 발생하여 일관성이 조금 떨어집니다.

**Page 12**

* 다음은 NIR2VC입니다.
* Pix2pix는 VNIR 결과보다 심한 영상 왜곡이 발생합니다. Pix2pixHD는 신호등 색이 잘못 생성되는 오류가 발생합니다. Baseline은 수치상으로 가장 높지만 시각적으로 색의 일관성이 매우 떨어집니다.
* VAE 적용 방법은 좋은 texture, 일관성이 보이지만 overfitting 결과 발생합니다.
* VQ-VAE 적용 방법은 영상이 깨지는 듯한 artifact가 생성됩니다.
* 마지막 방법은 baseline에 비해 색의 일관성과 나무의 texture 등이 훨씬 개선된 것이 확인됩니다.
* 요약하면, VAE와 VQ-VAE를 적용한 모델은 기대와는 다르게 성능 개선은 어려웠습니다. 이는 VAE와 VQ-VAE는 근적외선 채색 task에 맞지 않고 오히려 학습에 방해를 초래하여 성능 하락과 artifact 생성의 원인이 됩니다.

**Page 13**

* Edge와 feature fusion module을 적용한 방법이 baseline보다 정량적/정성적으로 개선되어, 4개의 지점에서 결과를 확대해 추가비교해보겠습니다.
* Position 0와 position 3에서 baseline과 비교하여, 건물 외벽에서 일관성을 확인할 수 있고, texture가 baseline보다 뛰어납니다. 특히 position 3에서는 사람 얼굴이 약간 생성됩니다.

**Page 14**

* position 1과 2에서도 역시 texture 생성능력이 좋으며, 특히 표지판이 더 선명하게 생성됩니다. 그러나, 오른쪽의 도로, 왼쪽의 자동차 같은 회색을 추정하는 능력은 다소 떨어지고, 빨간색과 파란색 artifact가 발생합니다.
* 요약하면, edge와 feature fusion을 적용한 방법은 baseline보다 texture와 색의 일관성 면에서 뛰어나지만, 회색 같은 특정 색을 추정하는 능력은 떨어지고, 약간의 artifact가 발생합니다. 그리고, 전체적으로 영상이 조금 밝게 생성됩니다.

**Page 15**

* 다음은 ablation study 중 먼저 canny filter에 대한 결과입니다.
* Canny filter는 가우시안 필터와 소벨 필터가 사용되며, 예를 들어, 왼쪽 그림처럼 가우시안 커널은 고정하고, 소벨 커널을 증가할수록 더 다양한 엣지가 생성되고 작을수록 디테일한 엣지가 생성됩니다.
* 또한 canny filter는 오른쪽 그림처럼 비최대억제를 통해 두꺼운 엣지를 얇은 엣지로 변환할 수도 있는 알고리즘입니다.

**Page 16**

* 저는 먼저 두꺼운 엣지중 소벨 커널을 다르게 한 결과를 비교하고 두꺼운 엣지와 얇은 엣지를 사용할 때 결과를 비교했습니다.
* 먼저 두꺼운 엣지 중 커널이 3일 때 성능이 더 높게 측정됩니다. 이는 커널 크기가 클수록 더 다양한 엣지가 추출되지만 불필요한 정보도 함께 포함되기 때문입니다.
* 또한, 두꺼운 엣지가 얇은 엣지보다 성능이 더 좋았습니다. 얇은 엣지는 두꺼운 엣지보다 디테일 정보가 부족한 것으로 보입니다.
* 결과 영상을 보시면 커널이 3이고 두꺼운 엣지를 사용할 때, 텍스처와 일관성이 좋은 것을 확인할 수 있습니다.

**Page 17**

* 다음으로 feature fusion module에 대한 ablation study를 진행했습니다.
* 총 4가지의 fusion 방식을 비교하였으며, 오른쪽 그림의 AFF는 MS-CAM을 기반으로 한 방법으로, 출력 채널이 입력 채널의 절반으로 생성됩니다. 즉, 3채널의 2가지 입력을 사용하여 3채널의 출력을 생성하고, MS-CAM과 제안 방법은 6채널의 출력을 생성합니다.
* 저는 채널을 압축하는 방식보다 그대로 유지해 정보 보존하고, L과 ab의 서로 다른 정보를 독립적으로 학습하는 것이 더 좋을 것이라 예상했습니다.
* 표를 보시면 MS-CAM과 제안하는 MS-CAM 모두 성능이 향상되고, 특히 MS-CAM은 VNIR2VC에서, 제안방법은 NIR2VC에서 강점을 보입니다. 그러나, AFF는 일반 concatenation보다 오히려 성능이 떨어지는 모습을 보입니다.
* 이는 AFF의 3채널로 압축되는 feature fusion 방식은 그전까지 학습했던 정보들이 사라지는 것을 의미합니다. 또한, 정보를 압축하는 것보다는 학습을 통해 융합하는 것이 성능을 더 향상시킬 수 있으며, 추가로 L과 ab의 밝기와 색차라는 정보는 서로 독립적으로 융합하는 것이 더 좋은 것을 의미합니다.
* 출력 영상에서는 MS-CAM을 사용할 때, 왼쪽 상단에 주황색 artifact가 발생하였으며, 학습에 적합하지 않음을 알 수 있습니다.
* 요약하면, 예상대로 채널 수를 유지하고 서로 다른 정보는 독립적으로 학습하는 것이 성능에 더 좋은 영향을 주었습니다.

**Page 18**

* 마지막으로 제안 방법에 GAN을 추가하였을 때를 비교했습니다.
* GAN은 mode collapsing 문제를 해결하기 위해 분포를 고르게 학습하고, overfitting을 감소시킵니다.
* 저는 GAN이 artifact와 PSNR을 개선할 거라 생각했습니다.
* 표를 보시면, WGAN을 사용하였을 때, VNIR2VC의 PSNR만 개선된 것을 확인했습니다.
* 출력 영상에서는 GAN 및 WGAN을 사용할 때, 일관성이 떨어지는 것을 확인했습니다. 그러나, WGAN을 적용한 방법은 기존 방법에서 생성되지 않던 다채로운 색을 생성하였으며, 표지판의 파란색이 조금 생성되는 것을 확인했습니다.
* 따라서, 예상과는 반대로 GAN 및 WGAN은 오히려 성능이 떨어지는 결과가 발생하지만, WGAN은 최적화를 통해 성능 개선의 가능성을 확인했습니다.

**Page 19**

* 마지막 결론입니다.
* 근적외선 영상 채색 기법에 대해 기존 방법을 개선하기 위해 처음에 VAE와 VQ-VAE의 특성을 활용하고자 했습니다. 그러나, 기대와 달리 VAE의 특성은 성능 하락과 artifact를 생성하는 등 colorization task에 맞지 않음을 보여주었습니다.
* 따라서 방향을 바꿔 detail과 일관성이 떨어지는 기존 방법의 문제를 개선하고자 edge detction과 feature fusion module을 사용했습니다.
* 그 결과, 수치적으로 NIR2VC의 PSNR을 약 0.8 개선하였고, 영상에서 색의 일관성과 texture 생성 능력을 개선했습니다.
* 향후 연구로는 artifact가 생성되는 문제 해결, WGAN을 이용한 성능 개선, 그리고 모델 경량화를 수행하고자 합니다.