# 라이트 필드 데이터를 이용한 학습 기반 다중 초점 이미지 합성 신경망

김형식 O, 김영섭
단국대학교 전자전기공학부
단국대학교 전자전기공학부
32181265@dankook.ac.kr, wangcho@dankook.ac.kr

## 요 약

본 논문에서는 라이트 필드 데이터를 이용한 다중 초점 이미지 합성 신경망 LFFCNN(Light Field Focus Convolutional Neural Network)을 제시한다. 다양한 초점의 이미지를 학습하기 위하여 각기 다른 초점의 이미지들과 한 장의 GT 이미지를 한 쌍으로 갖는 라이트 필드 데이터를 사용하여 신경망을 학습하였다. 입력 이미지는 특징 추출 모듈을 통해 초점 정보가 특징맵으로써 추출된다. 그 후 해당 초점 정보를 합성하기 위해 요소별 최대값(Elementwise Maximum)을 반환하는 방식으로 모든 이미지의 초점정보가 통합된 특징맵을 얻고 이후 특징 재구성 모듈을 통해 특징맵을 이미지로 복원하게 된다. 실험에 쓰인 테스트 데이터셋 역시 라이트 필드 데이터를 사용하였고 LFFCNN 을 사용한 결과 라이트 필드 데이터셋과 같이 다양한 수의 다중 초점 이미지를 합성하더라도 초점 이미지를 적절하게 합성하는 것을 확인할 수 있다.

## 1. 서론

다중 초점 이미지 합성의 목표는 다양한 초점을 가진 이미지들을 한장의 모든 깊이에 대한 초점을 가진 올인포커스(All-in-Focus) 이미지를 합성하는 것이다. 이러한 합성 방식은 크게 두 가지로 나뉜다. 결정맵을 출력으로 갖는 공간 영역 방식과 올인포커스 이미지를 출력으로 갖는 종단간 신경망 방식의 변환 영역 방식으로 나눌 수 있다[1]. 결정맵 방식의 경우 예를 들어 입력 이미지가 두 장의 초점 이미지라면 신경망의 출력인 결정맵이 초점이 잡힌 곳을 마스킹해주는 이진 이미지이므로 결정맵을 입력 이미지와 곱해주는 후처리 과정이 필요하다는 것이 단점이다. 변환 영역 방식의 경우 입력 이미지와 출력 이미지가 직접적으로 신경망을 통해 연결되기 때문에 후처리 과정이 필요 없다.

많은 종단간 신경망의 경우 인코더 디코더 구조를 사용한다[2]. 인코더를 통해 점점 저차원으로 차원을 축소하고 채널을 늘려가며 모든 곳의 초점 정보를 가진 특징맵을 획득하고 디코더가 저차원 특징맵을 다시 고차원 이미지로 복원한다. 하지만 이러한 방식의 경우 인코더의 차원 축소가 공간적 정보의 손실을 야기하기 때문에 디코딩을 통한 원본 해상도를 복원한다 하더라도 공간 정보 손실이 생긴다. 이를 방지하기 위해 IFCNN[3]는 신경망의 모든 구간에서 특징맵의 해상도를 유지하는 합성곱 신경망을 사용했다. 본 논문에서는 공간 정보의 손실을 방지하기 위해 고해상도와 저해상도의 특징을 모두 포괄할 수 있는 다중 초점 이미지 합성 신경망 LFFCNN(Light Field Focus Convolutional Neural Network)을 제안한다.

본 논문은 2 장에서 본 신경망의 구조를 설명하고 3 장은 LFFCNN 의 실험 결과, 4 장에서 결론을 낸다.

## 2. 신경망 구조

### 2.1 라이트 필드 데이터셋

라이트 필드 데이터셋은 LFDOF[4]를 사용한다. 일반적인 라이트 필드 데이터셋은 다중 초점에 대한 이미지가 없다. 라이트 필드 이미지는 라이트 필드 카메라의 특성상 재초점이 가능하므로 기존 데이터셋의 이미지를 재초점화 하는 과정을 통해 다중 초점 이미지로 만들어 데이터셋을 구성한다. 재초점화 하여 얻은 데이터 한 쌍은 다음과 같이 구성된다.

1. 입력:각기 다른 초점을 가진 이미지 6~15 장
2. 라벨:모든 곳에 초점이 맞은 올인포커스(All-in-Focus)이미지 한 장

### 2.2 신경망 구성

본 이미지 처리 분야에서 합성곱 신경망은 이미지의 특징을 추출하는 역할을 하며 이를 이용한 FCN(Fully Convolutional Network)은 이미지 처리에서 전통적인 변환 영역 방식과 유사하다. 본 신경망은 FCN 구조로써 종단간 학습을 진행하며 크게 세 모듈로 나눌 수 있다. 입력 이미지의 초점 정보를 추

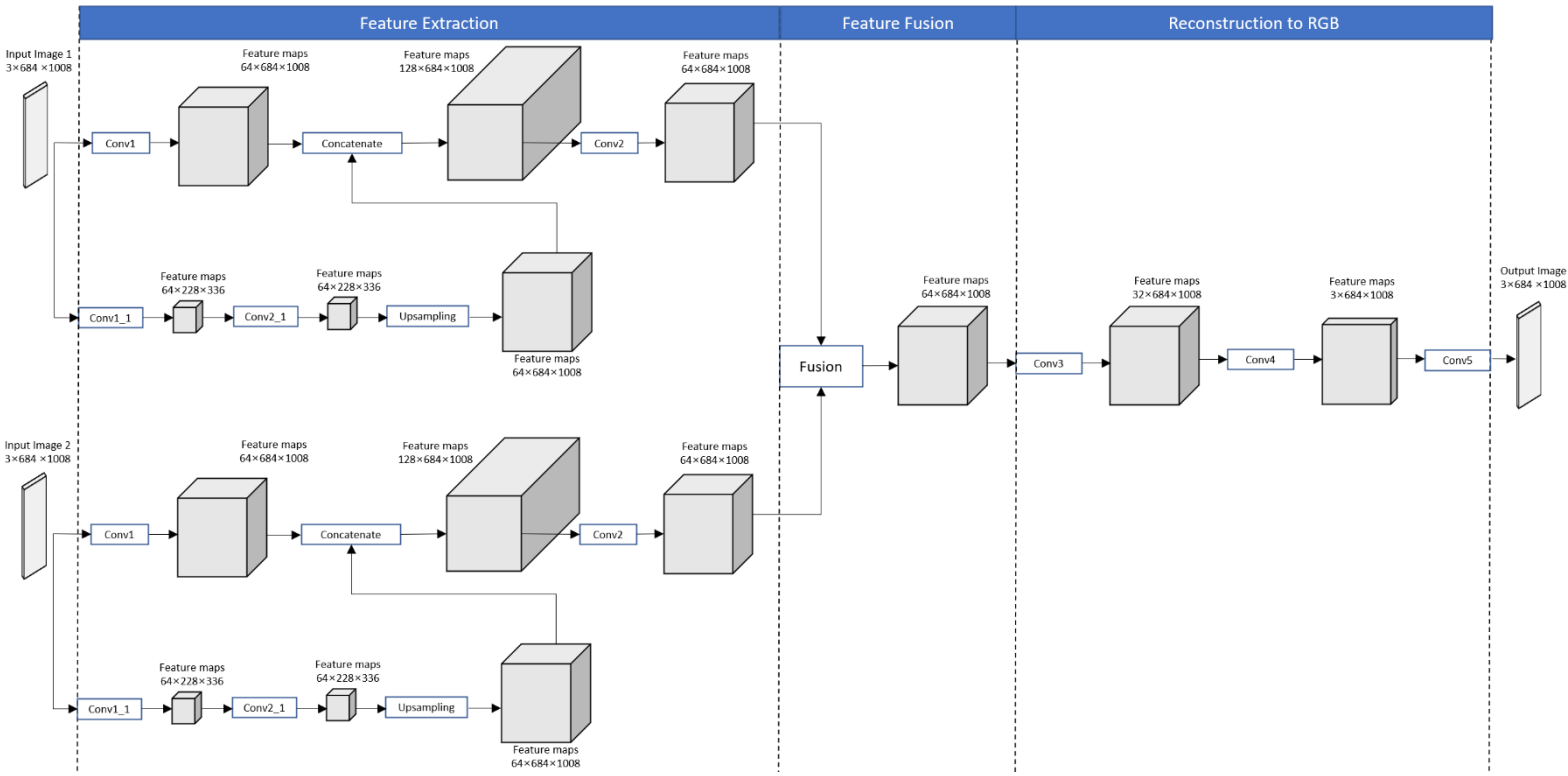


**그림 1 신경망 구조(2개의 초점 이미지 예시)**

출하는 특징 추출 모듈, 각 이미지들의 초점 정보를 하나의 특징맵으로 모으기 위해 합성을 진행하는 특징 합성 모듈 그리고 마지막으로 모든 초점 정보를 가진 특징맵을 이미지로 복원하는 특징 재구성 모듈이 있다.

### 2.2.1 특징 추출 모듈

특징 추출 모듈은 기존의 인코더 구조의 역할과 동일하게 입력 이미지의 초점 정보를 특징맵으로 추출한다. 하지만 특징맵의 해상도를 줄인다면 공간적 정보의 손실이 있을 수 있기 때문에 본 신경망에서는 그림 1 과 같이 하나의 입력 이미지에 대해 고해상도 특징맵과 저해상도 특징맵으로 나누어 특징 추출을 진행한다. 합성곱 신경망에서 이미지 회귀 문제는 안정적인 학습이 어렵다. 이는 이미지 분류 문제에서 사전 학습된 합성곱의 가중치를 사용하여 해결할 수 있다. 따라서 본 모듈의 Conv1, Conv1_1 은 ResNet101 의 첫 번째 합성곱 가중치를 사용한다.

입력 이미지는 Conv1 과 Conv1_1 을 거치며 고해상도 특징맵과 저해상도 특징맵 두 가지로 특징 추출을 한다. Conv1_1 에 의해 저해상도 특징맵은 본 이미지 해상도 대비 크기가 1/3 이 감소한다. 합성곱을 지난 이후 쌍선형 보간법을 통해 원래 해상도로 복원되어 같은 크기의 두 특징맵을 합성하며 고해상도와 저해상도 특징맵의 정보를 모두 포함하는 특징맵을 만든다. 합성은 두 특징맵을 채널 방향으로 쌓고(Concatenate) 합성곱(Conv2)을 거치는 방식으로 진행되며 늘어난 특징맵의 채널 수를 Conv2 를 통해 줄이며 두 특징맵 정보 합성을 진행한다.

### 2.2.2 특징 합성 모듈

특징 합성 모듈은 추출된 초점 정보를 가진 특징맵들을 합성한다. 이 과정을 통해 각 특징맵에서 초점이 잡힌 영역만을 추출하여 하나의 특징맵을 만드는데 그림 2 와 같이 초점 이미지에서 초점이 잡힌 곳의 픽셀값은 다른 영역보다 상대적으로 높은 값을 갖는다. 따라서 다양한 초점 이미지의 특징맵을 최대값을 반환하는 최대값 반환(Elementwise-maximum) 함수를 사용한다.

해당 모듈의 합성은 특징 추출에서의 합성과 다르게 요소별 비교를 통해 최대값을 반환하는 방법을 사용한다. 기존 합성을 사용하지 않는 이유는 합성곱은 입력 채널과 출력 채널을 초매개변수로써 지정해주어야 하는 관계로 특정 수의 입력만 합성할 수 있다. 또한 라이트 필드 이미지의 특성상 입력 다중 초점 이미지의 수가 확정되어있지 않으므로 다양한 수의 입력에 대응할 수 없다. 반면 최대값 반환의 경우 입력 이미지의 요소별 픽셀을 비교하므로 매개변수를 가질 필요가 없어 다양한 수의 입력을 합성할 수 있기 때문에 해당 방법을 사용한다.

### 2.2.3 특징 재구성 모듈

특징 재구성모듈은 인코더 디코더 형식의 다른 신경망에서 디코더와 같이 필요한 초점 정보를 한데 모은 특징맵을 3 차원 RGB 이미지로 복원하는

역할을 한다. 특징 합성 모듈에서 얻어진 결과물은 이미지가 아닌 채널의 수가 64 개인 특징맵이기 때문에 재구성 모듈의 합성곱을 통해 차원의 수를 줄이며 이미지를 복원한다.

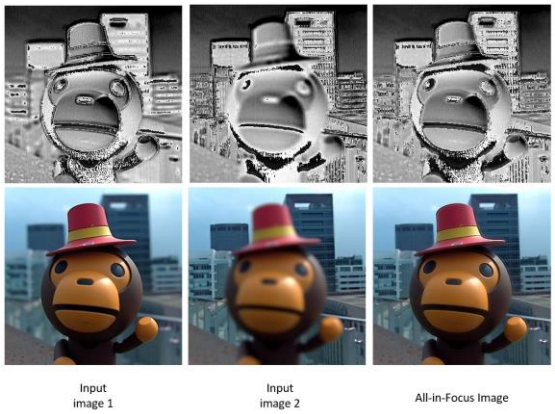


**그림 2 이미지 특징맵 시각화**

## 3. 실험 결과 및 분석

실험 장비로는 CPU intel Xeon-Silver 4210R, GPU NVIDIA Quadro RTX 6000, RAM 64GB 를 사용했으며 소프트웨어로는 우분투 20.04 에서 파이토치를 이용하여 실험을 진행하였다.

데이터셋은 앞서 기술했듯이 라이트 필드 데이터의 재초점 이미지를 사용하였다. 훈련용 데이터 890 쌍의 이미지를 데이터 증강 기법을 사용하여 약 5300 장의 이미지 쌍으로 신경망을 학습하고 50 장의 테스트 이미지를 사용하여 평가를 진행하였다.

데이터셋에 포함된 이미지는 688×1008×4 의 크기를 갖는다. 이미지의 차원은 순서대로 RGBD 이므로 마지막 D(Depth map)를 제외한 3 차원만 사용하여 최종적으로 사용되는 이미지는 688×1008×3 이다.

수식 1,2,3 은 순서대로 평가 지표 PSNR 과 PSNR 에 포함된 MSE, 그리고 SSIM 을 나타낸다. 그림 3 은 본 신경망을 통해 합성된 이미지와 라벨 이미지를 비교한 것이다.

$$PSNR = 10\,log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad \text{(수식1)}$$

$$MSE = \frac{1}{m\,n}\sum_{m}\sum_{n}[I(i,j) - K(i,j)]^2 \quad \text{(수식2)}$$

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \quad \text{(수식3)}$$

표 1. 신경망 성능 분석

| 평가 지표 | PSNR | SSIM |
|---|---|---|
| 점수 | 24 | 0.85 |


**그림 3 신경망 결과물 비교**

## 4. 결론

본 논문은 라이트 필드 데이터를 이용하여 종단간 학습방식의 다중 초점 이미지 합성 신경망을 제시했다. 학습에 사용된 라이트 필드 이미지는 라이트 필드 이미지를 재초점화하여 만든 데이터로 다중 초점 이미지와 올인포커스 이미지가 하나의 쌍을 이루어 학습에 사용된다. 실험을 통해 학습된 신경망이 각 이미지의 초점정보를 합성하여 하나의 올인포커스 이미지를 출력하는 것을 볼 수 있다. 결과물을 육안으로 확인해보았을 때 본 신경망의 목표인 라이트 필드 데이터셋을 사용하여 다양한 수의 다중 초점 이미지를 합성한 결과물이 목표 이미

지와 유사하게 초점 정보를 잘 습득한 모습을 확인할 수 있다. 또한 사용된 데이터셋의 절대적인 수가 부족한 것을 감안할 때 추후 데이터의 추가와 구조 개선을 통해 제시한 신경망을 고도화할 수 있을 것으로 기대된다.

## 감사의 글

NRF-2020R1A2C2008717 의 지원으로 본 논문을 제출합니다.

## 참고문헌

[1] Zhang, Xingchen. "Deep learning-based multi-focus image fusion: A survey and a comparative study." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[2] Pan, Tao, et al. 'A novel multi-focus image fusion network with U-shape structure." Sensors 20.14 (2020): 3901

[3] Zhang, Yu, et al. "IFCNN: A general image fusion framework based on convolutional neural network." *Information Fusion* 54 (2020): 99-118.

[4] Ruan, Lingyan, et al. "Aifnet: All-in-focus image restoration network using a light field-based dataset." *IEEE Transactions on Computational Imaging* 7 (2021): 675-688.