

태양광 패널 결함 탐지 및 분류

IIE 4123 딥러닝과응용 Term Project

Table of Contents

1	서론	3
1.1	연구 배경 및 동기	3
1.2	선행 연구 조사	4
A	결함 탐지 메커니즘 선정 배경	4
B	후보 모델 선정 배경	4
1.3	프로젝트 목적 및 필요성	5
2	본론	7
2.1	데이터 소개 및 전처리	7
A	데이터 소개	7
B	데이터 전처리	7
2.2	모델 비교 및 선정	8
A	VGG-19	8
B	MobileNet-V3	9
C	EfficientNet-b0	9
D	DenseNet-121	9
E	최종 모델: EfficientNet-b0	10
2.3	모델 성능 개선	10
A	데이터 증강	10
B	Fine Tuning	11
C	Drop out	11
D	Class 불균형	11
E	하이퍼파라미터	12
3	결론	14
3.1	결과 해석 및 시각화	14
3.2	한계점	14
3.3	개선 방안	15
3.4	기대 효과 및 프로젝트 의의	16
4	Reference	18

1 서론

1.1 연구 배경 및 동기

최근 환경 오염, 기후 변화, 그리고 자원 고갈에 대한 우려가 커지면서 차세대 친환경 에너지 개발의 필요성이 그 어느 때보다 중요하게 부각되고 있습니다. 다양한 친환경 에너지원 중에서도 태양광 에너지는 지속 가능하며 뛰어난 확장성을 지닌 에너지원으로서 가장 높은 잠재력을 지닌 대안으로 주목받고 있습니다. 특히, 전력 생산 과정에서 탄소 배출이 거의 없다는 점은 환경적 부담을 획기적으로 줄이는 데 기여하며, 사실상 무한으로 공급 가능한 태양빛을 에너지원으로 활용한다는 점에서 태양광 에너지의 가치는 더욱 부각됩니다. 이와 더불어 전 세계적으로 탄소 중립을 목표로 한 움직임이 빠르게 확산되면서 태양광 에너지는 차세대 핵심 에너지원으로서 그 입지를 더욱 견고히 다지고 있습니다. 국제에너지기구(IEA)에 따르면, 향후 10 년간 태양광 에너지 시장이 지속적으로 성장할 것으로 전망됩니다.¹ 이러한 성장세에 힘입어, 친환경 방식으로 생산되는 전체 에너지 중 태양광 발전의 비중이 가장 커질 것으로 전망며, 전 세계 에너지 전환의 핵심 동력으로서 태양에너지가 중추적인 역할을 하게 될 것으로 기대됩니다.²

그러나 태양광 에너지는 이처럼 뛰어난 잠재력과 장점에도 불구하고 여전히 극복해야 할 한계가 존재합니다. 현재로서는 태양광 에너지의 발전 효율과 경제성을 고려할 때, 태양광 에너지만으로 에너지 생산을 전적으로 의존하기에는 부족한 점이 있습니다. 특히, 이론적으로 규명된 발전 효율의 한계와 다른 친환경 에너지원에 비해 상대적으로 높은 발전 단가는 태양광 에너지의 상용화를 가로막는 주요 요인으로 작용하고 있습니다.³ 지난 30 여 년간 이러한 한계를 극복하기 위해 재료 및 발전 원리에 대한 다양한 연구가 진행되어 왔지만, 실제 시스템에서 달성된 최대 효율은 약 40%에 불과하며, 성능 향상에 있어 뚜렷한 진전이 이루어지지 않았습니다.⁴⁵ 이러한 정체된 연구 결과는 기존의 재료 및 발전 원리 중심의 전통적인 접근 방식에서 벗어나, 새로운 관점에서 문제를 해결해야 할 패러다임 전환의 필요성을 강조합니다.

한 가지 합리적인 접근법은 태양광 에너지의 주요 특징 중 하나인 발전량이 외부 요인에 크게 영향을 받는다는 점에 주목하는 것입니다. 즉, 태양광 발전량은 태양광 패널 표면에 도달하여 흡수되는 빛(광자)의 양에 따라 달라지므로, 패널 표면에서 최적의 태양광 노출을 확보하는 것이 발전량을 증가시키는 데 매우 중요합니다. 이를 위해서는 패널 표면의 상태를 정확하게 모니터링하고 탐지하는 기술이 필요한데, 이는 딥러닝이 뛰어난 역량과 강점을 보이는 분야 중 하나입니다. 따라서 본 프로젝트에서는 태양광 패널 표면 이미지를 활용해 표면 결함을 탐지함과 동시에 결함 유형을 분류할 수 있는 딥러닝 기반의 새로운 프레임워크를 구축하고자 합니다.

1.2 선행 연구 조사

A 결함 탐지 메커니즘 선정 배경

본 프로젝트를 진행하기에 앞서 '태양광 패널 표면 결함 탐지 및 분류'라는 주제와 관련하여 유사한 연구가 이전에 진행된 적이 있는지 조사해보았습니다. 이를 통해 기존 연구가 있다면 어떠한 방식으로 진행되었는지, 그리고 여전히 해결되지 않은 연구 공백(gap)은 무엇인지 파악한 후 본 프로젝트에서는 이를 해결하기 위한 방향성을 정립하고자 선행 연구를 검토하였습니다. 조사 결과, 본 프로젝트와 유사한 주제인 '태양광 패널 이상 탐지'와 관련된 연구는 일부 존재했지만, 대부분 예측된 발전량 데이터, 날씨 데이터, 인버터 신호 데이터 등을 기반으로 이상을 탐지하는 연구가 대부분인 것으로 확인되었습니다.⁶

이처럼 현재 대중적으로 사용되는 태양광 패널 결함 탐지 메커니즘은 크게 두 가지로 분류할 수 있습니다. 하나는 예측된 발전량 데이터를 활용하여 결함을 탐지하는 방식이고, 다른 하나는 패널 내부 인버터에서 발생하는 이상 신호를 활용하여 결함을 탐지하는 방식입니다. 전자의 경우 태양광 패널을 통해 실제로 생산된 전력량이 예측된 전력 생산량보다 낮을 경우 결함이 있다고 판단하는 방식입니다.⁷ 이렇게 결함이 있을 것으로 식별된 패널은 이후 점검을 통해 결함 여부를 확인하고, 그에 따른 적절한 조치가 취해집니다. 이와 같은 방법은 기본적인 결함의 탐지가 가능하기는 하지만, 결함의 유형을 정확히 분류하고 분석하는 데에는 한계가 있으며, 현장 점검이 필수적으로 동반되어야 한다는 점에서 비효율적입니다. 또한, 낮은 전력 생산량이 반드시 패널의 결함으로 인한 것이 아니라 기상 악화 등으로 태양 빛이 패널에 충분히 도달하지 못하는 외부 요인으로 인해서도 발생했을 수 있기에, 정확한 결함 진단이 어려우며, 해당 탐지 방법의 모호성을 부각합니다. 후자의 경우 태양광 패널 내부의 인버터에서 이상 신호가 감지되면 결함이 있다고 판단하고, 해당 패널을 점검하여 적절한 조치를 취하는 방식입니다.⁸ 해당 방식은 실제 이상 신호를 활용하여 탐지하기에 앞서 언급한 첫 번째 방식보다는 더욱 정확한 탐지가 가능하지만, 인버터에서 이상신호가 감지되는 시점에는 이미 내부적으로 상당한 문제가 발생했을 가능성이 높습니다. 이와 같은 탐지 시점의 지연은 직렬로 연결된 다른 패널에도 부정적인 영향을 미쳐, 시스템 전체의 큰 문제로 확산될 우려가 있습니다.

이와 더불어, 태양광 발전 효율 저하의 원인은 태양광 패널 내부의 회로나 시스템의 결함과는 직접적인 관련이 없는 경우가 많으며, 대부분 경미한 결함으로부터 기인한다는 연구 결과가 확인되었습니다.⁹ 이러한 경미한 결함의 요인으로는 먼지, 오염물질 등이 패널을 가려 태양빛의 입사를 방해하거나, 패널 표면의 가벼운 물리적 손상으로 태양광이 입사를 못하는 등 패널 표면과 관련된 문제가 대부분을 차지합니다. 그러나 현재 널리 사용되고 있는 탐지 메커니즘으로는 발생 빈도가 높은 표면 관련 문제를 효과적으로 탐지하는 데 한계가 있습니다. 아울러, 최근 태양광 에너지가 차세대 주요 친환경 에너지로 주목받으면서 사막과 같은 외진 지역에 대규모 발전소들이 잇따라 건설되고 있습니다.¹⁰ 이러한 발전소의 관리를 위해 앞서 언급한 두 가지 탐지

방법을 적용할 경우, 기존 방법들의 한계점으로 인해 상당한 기회비용이 발생할 것으로 예상됩니다. 이와 같은 한계점들을 종합적으로 고려했을 때, 태양광 패널 표면을 확인하여 빠르고 정확하게 결함을 탐지하고 결함의 유형을 분류할 수 있는 딥러닝 모델의 필요성이 더욱 명확해졌습니다. 따라서, 이러한 작업을 효과적으로 수행할 수 있는 딥러닝 모델을 도입하면 발전 효율과 안정성을 동시에 향상시켜 전체적인 발전 비용을 절감할 수 있을 뿐만 아니라, 궁극적으로 태양광 에너지의 상용화를 가속화하는 데 중요한 역할을 할 것으로 기대됩니다.

B 후보 모델 선정 배경

Computer Vision 분야에서 지금까지 다양한 작업을 수행하기 위한 여러 알고리즘과 딥러닝 모델이 연구되어 왔습니다. 대표적으로, Image processing 분야에서 획기적인 전환점을 마련한 CNN (Convolutional Neural Network) 아키텍처 기반의 모델부터, 이상 탐지에 특화된 Auto Encoder (AE)와 Generative Adversarial Network (GAN), 그리고 주변 맥락을 파악하여 더욱 복잡한 문제에 대한 해결책을 제공하는 Vision Transformer (ViT)와 같은 Transformer 기반 모델에 이르기까지 폭넓은 모델들이 연구되어 사용되고 있습니다.¹¹¹²¹³¹⁴ 이러한 모델들은 초기 모델보다는 이를 발전시키고 부족한 부분을 보완한 후속 모델에서 더욱 높은 정확도를 보이는 것이 일반적이지만, 간과해서는 안 되는 중요한 점은 바로 해당 모델이 수행할 특정 작업에 적합한지의 여부입니다. 예컨대, 높은 정확도를 보이는 모델이라 하더라도, 학습에 방대한 양의 데이터가 필요해 특정 환경에서는 도입이 어려운 경우가 있을 수 있습니다. 또한, 구조가 지나치게 복잡한 모델은 연산 과정에서 많은 시간이 소요되어 실시간 처리가 중요한 작업에는 적합하지 않을 수 있습니다. 이러한 모델의 특성은 작업의 요구사항에 따라 신중히 고려되어야 합니다.

본 프로젝트에서는 태양광 패널의 결함 탐지 및 분류를 목표로 하며, 해당 작업의 주요 요구사항으로는 빠르고 정확하면서도 안정적인 성능을 제공해야 한다는 점, 그리고 모델이 임베디드 시스템에 탑재될 가능성을 고려해야 한다는 점이 있습니다.¹⁵ 따라서 대규모 데이터에서 좋은 성능을 보이는 ViT 나, 분류보다는 이상탐지 작업 수행에 초점을 맞춘 AE 와 GAN 모델 보다는 상대적으로 가볍고, 결함을 탐지하면서 동시에 유형별 분류 작업도 잘 수행할 수 있는 CNN 아키텍처 기반의 모델 내에서 후보 모델을 선정하는 것이 적합하다고 판단했습니다. 그리고 여러 종류의 CNN 아키텍처 기반의 모델들을 평가하기 위해 고려해야 할 주요 지표로는 모델의 파라미터 수, 연산 처리 효율성, 네트워크 구조의 간결성, 그리고 견고한 성능을 확보하기 위해 필요한 학습 데이터의 양 등이 있습니다. 본 프로젝트에서는 이러한 지표들을 바탕으로 VGG-16, MobileNet-V3, EfficientNet-b0, DenseNet-121 총 4 개의 모델로 후보로 추렸습니다.

1.3 프로젝트 목적 및 필요성

앞서 분석한 바와 같이 태양광 에너지는 환경 친화적이고, 지속 가능하며, 확장성 또한 우수한 덕분에 차세대 핵심 에너지원으로 주목받고 있는 한편, 몇 가지 한계점들로 인해 경제적 실효성과 운영 효율성 측면에서는 여전히 다소 부족한 부분이 존재한다는 사실을 알 수 있습니다. 특히, 태양광 패널의 결함으로 인한 발전 효율 저하와 이를 탐지하고 해결하는 과정에서 발생하는 많은 시간적·경제적 비용은 이러한 한계의 주요 원인으로 지적되고 있습니다.

현재 사용되는 결함 탐지 메커니즘은 결함 유형의 불명확성과 탐지 시점의 지연성, 직접적 확인이 동반되는 수동성으로 인해 다소 비효율적입니다. 더 나아가, 기존 시스템은 빈번하게 발생하는 태양광 패널 표면에서 발생하는 여러 가지 결함을 효과적으로 탐지하지 못하고 있으며, 이는 발전 효율 저하로 이어지는 중요한 문제입니다. 그리고 최근 태양광 발전소가 사막과 같이 인간의 접근이 어려운 지역에 대규모로 잇따라 건설됨에 따라 이러한 한계가 더욱 부각되고 있습니다.

이에 따라, 표면에서 발생하는 결함을 신속하고 정확하게 탐지하며, 탐지된 결함의 유형을 체계적으로 분류할 수 있는 시스템의 필요성이 더욱 부각되고 있습니다. 딥러닝 기반의 결함 탐지 및 분류 기술은 이러한 문제를 효과적으로 해결할 수 있는 대안으로, 태양광 패널의 표면 상태를 정밀하게 분석하여 효율적인 관리 방안을 제공할 수 있을 것으로 기대됩니다. 본 프로젝트는 이러한 필요성을 충족하는 최적의 딥러닝 모델을 개발하여 태양광 에너지와 관련된 주요 한계들을 해결하고자 합니다.

구체적으로, 본 프로젝트는 다음과 같은 세부적인 목표를 달성하는 것을 목적으로 합니다. 먼저, 여러 CNN 아키텍처 기반 모델의 구조와 특징을 비교·분석하여 태양광 패널 결함 탐지 및 분류 작업에 가장 적합한 후보 모델을 선정할 것입니다. 이후, 각 모델의 성능을 실험적으로 비교하여 높은 정확도 뿐만 아니라 학습 완료에 소요되는 시간(training time)과 추론 속도(inference speed)까지 종합적으로 고려해 최적의 모델을 선정할 것입니다. 이때, 모델들의 성능을 동일한 실험 환경 (동일한 dataset 과 hyper-parameter)에서 평가할 수 있도록 주의를 기울여 평가할 것입니다. 최적 모델이 선정되면, 적절한 데이터 증강 기법과 하이퍼 파라미터 등 모델의 성능에 영향을 미치는 다양한 요소를 조정하며, 최고의 성능(정확성, 안정성, 효율성)을 발휘할 수 있도록 고도화 작업을 진행할 것입니다. 이와 같은 세부 목표를 단계적으로 달성함으로써 최적의 딥러닝 모델을 설계하고, 궁극적으로 태양광 에너지 기술의 발전에 기여할 수 있는 혁신적인 기술을 구현하고자 합니다.

2 본론

2.1 데이터 소개 및 전처리

A 데이터 소개

본 프로젝트에서는 태양광 패널의 정상 상태와 다양한 결함 상태를 포함한 총 **875개**의 이미지를 학습 데이터로 사용하였습니다. 데이터는 Kaggle 에서 공개된 데이터셋을 활용하였으며, 다양한 환경에서 촬영된 태양광 패널의 이미지를 포함하고 있습니다.

데이터는 총 **6 개의 클래스**로 분류되었으며, 각 클래스는 태양광 패널의 상태를 반영합니다. 각 클래스의 이름과 클래스 당 데이터 개수, 그리고 클래스별 특징은 다음과 같습니다.

정상 상태

- **Clean** (194 개): 패널 표면에 결함이나 오염이 없는 상태.

결함 상태

- **Bird-drop** (192 개): 조류 배설물이 패널 표면에 쌓여있는 상태.
- **Dusty** (191 개): 먼지로 인해 패널이 더럽혀진 상태.
- **Electrical-Damage** (104 개): 과전압 등으로 인한 전기적 손상 상태.
- **Physical-Damage** (70 개): 외부 충격이나 날씨로 인해 패널 표면에 균열이나 파손이 발생한 상태.
- **Snow-Covered** (124 개): 패널 표면이 눈으로 덮여 태양광 흡수가 방해된 상태.

이 데이터는 결함 탐지를 위해 중요한 시각적 특징을 포함하고 있습니다. 예를 들어, Bird-drop 클래스는 조류 배설물의 패턴을, Physical-Damage 는 균열 및 파손의 특징을, Snow-Covered 는 눈 덮임의 구조적 특징을 나타냅니다. 이를 통해 태양광 패널의 전력 효율 저하 원인을 보다 정확하게 파악할 수 있습니다.

다만, 데이터셋에는 클래스 불균형이라는 문제가 존재합니다. 예를 들어, 가장 많은 Clean 클래스는 194 개의 데이터를 보유한 반면, Physical-Damage 클래스는 70 개로 약 3 배의 차이가 발생합니다. 이러한 불균형은 모델이 특정 클래스에 편향되게 학습할 가능성을 높이며, 소수 클래스에 대한 탐지 성능 저하를 초래할 수 있습니다.

B 데이터 전처리

모델 학습의 성능을 극대화하고 데이터 품질을 높이기 위해 다음과 같은 전처리 작업을 수행하였습니다:

- 사진 영역 조정

데이터셋에는 태양광 패널 외부의 물체나 환경이 이미지의 상당 부분을 차지하는 사례가 존재했습니다. 예를 들어, 사람이 사진의 주요 영역을 차지하거나, 패널 주변 배경이 과도하게 포함된 경우가 있었습니다. 이러한 이미지는 태양광 패널 부분만 보이도록 ROI(Region of Interest)를 조정하거나 패널이 아예 잘 보이지 않는 경우 제거하여 데이터를 정제하였습니다. 이를 통해 불필요한 정보가 모델 학습에 영향을 미치지 않도록 하였습니다.

- 잘못된 라벨링 제거

데이터셋에 일부 잘못된 라벨링이 포함되어 있음을 확인하였습니다. 예를 들어, Physical-Damage 클래스에 속해야 할 데이터가 Bird-drop 클래스로 분류된 경우 등이 있었습니다. 이러한 데이터를 수동으로 검토하여, 올바르게 수정하거나 학습에 혼란을 줄 수 있는 애매한 데이터를 제거하였습니다.

이와 같은 전처리 과정을 통해 데이터셋의 품질을 개선하고, 모델 학습 과정에서의 신뢰성을 높일 수 있었습니다. 정제된 데이터셋은 모델의 성능 향상과 안정적인 학습에 기여하였음 test accuracy 의 평균과 분산을 통해 확인할 수 있었습니다. accuracy 평균은 80.4 에서 83.18 로 증가했고, 분산은 2.03 에서 1.733 으로 다소 감소했습니다.

2.2 모델 비교 및 선정

태양광 패널 결함 탐지 문제를 해결하기 위해 다양한 CNN 기반 모델을 실험적으로 비교하고, 최종적으로 최적의 모델을 선정하였습니다. 본 항목에서는 각 모델을 선정한 이유와 그 특징을 설명하며, 실험 결과를 기반으로 최종 모델 선정 과정을 서술하겠습니다.

모델을 비교하기 위해 정확도와 학습 시간이라는 두 가지 주요 지표를 사용하였습니다. 정확도는 모델이 결함 탐지를 얼마나 정확하게 수행할 수 있는지를 나타내는 핵심 성능 지표입니다. 태양광 패널의 결함 탐지는 발전 효율과 관련된 중요한 문제이기 때문에, 높은 정확도가 필수적입니다. 학습 시간은 모델을 실행할 때 걸리는 시간으로, 태양광 발전소의 특수성을 고려하여 중요한 평가 요소로 선정되었습니다. 예를 들어, 태양광 발전소는 드론과 같은 경량화된 장치에서 모델을 실시간으로 운영하는 것이 일반적이기 때문에, 빠른 학습 시간과 효율적인 연산이 필수적입니다. 이러한 지표를 바탕으로 모델들을 동일한 데이터셋과 환경에서 비교하여 최적의 모델을 선정하였습니다.

A VGG-19

VGG-19 는 19 개의 계층(16 개의 합성곱 계층과 3 개의 완전 연결 계층)으로 구성된 심층 신경망입니다.¹⁶ 이 모델은 3x3 필터를 사용하여 작은 영역의 세밀한 특징을 학습할 수 있으며,

균열, 변색, 오염과 같은 복잡한 결함 패턴을 학습하는 데 적합합니다. 이러한 이유로 높은 성능을 기대하며 채택하여 실험하였습니다.

<테스트 결과> 정확도(Accuracy): 74.30%, 학습 시간(Training Time): 약 15 분

VGG-19 는 비교적 높은 정확도를 보였으나, 연산량이 많아 학습 시간이 길어 실시간 모니터링과 같은 응용 사례에는 부적합한 한계가 있었습니다.

B MobileNet-V3

MobileNet-V3 는 경량화된 네트워크로, Inverted Bottleneck 구조를 활용하여 연산 효율성을 극대화한 모델입니다.¹⁷ 이 모델은 드론과 같은 경량 시스템에 탑재가 가능하며, 실시간 모니터링이 중요한 환경에 적합합니다. 특히, 연산량이 적어 상대적으로 빠른 학습 시간이 장점입니다.

<테스트 결과> 정확도(Accuracy): 75.42%, 학습 시간(Training Time): 약 8 분

MobileNet-V3 는 준수한 정확도와 빠른 학습 속도를 보였으나, 작은 모델 특성상 복잡한 결함 패턴을 학습하는 데 한계를 보였습니다.

C EfficientNet-b0

EfficientNet-b0 는 Compound Scaling 기법을 사용하여 네트워크의 Width, Depth, Resolution 을 균형 있게 조정한 모델입니다.¹⁸ 이 모델은 기존 CNN 모델과 달리, 모델 크기를 제한하지 않으면서도 높은 성능을 달성하도록 설계되었습니다. 또한 MobileNet 의 MBConv 블록을 활용하여 효율성을 극대화한 구조를 가지고 있습니다.

<테스트 결과> 정확도(Accuracy): 84.09%, 학습 시간(Training Time): 약 4 분

EfficientNet-b0 는 실험에 사용된 모델 중 가장 높은 정확도를 기록했으며, 매우 짧은 학습 시간을 통해 경량화된 모델의 강점을 입증하였습니다. 소규모 데이터 환경에서도 일반화 성능이 뛰어난 모델로 평가되었습니다.

D DenseNet-121

DenseNet-121 은 각 계층이 이전 계층의 출력을 참조하여 효율적으로 학습하는 네트워크 구조를 가지고 있습니다.¹⁹ 이러한 Dense 연결 구조 덕분에, 모델은 파라미터를 효율적으로 사용하며 결함 탐지와 같은 세밀한 특징 학습에 적합합니다.

<테스트 결과> 정확도(Accuracy): 82.68%, 학습 시간(Training Time): 약 15 분

DenseNet-121 은 높은 정확도를 보였으나, 학습 시간이 길어 실시간 처리에는 다소 부적합한 한계를 보였습니다.

E 최종 모델: EfficientNet-b0

위의 결과를 종합적으로 비교한 결과, EfficientNet-b0 를 최종 모델로 선정하였습니다.

EfficientNet-b0 는 실험한 모델 중 가장 높은 정확도인 84.09%를 기록했으며, 학습 시간은 약 4 분으로 매우 짧아 실시간 모니터링과 같은 응용 사례에 적합한 모델입니다. 또한 Compound Scaling 기법을 활용하여 네트워크의 균형적인 확장을 통해 소규모 데이터 환경에서도 일반화 성능이 뛰어납니다. 이와 더불어 EfficientNet-b0 는 파라미터 수와 계산량이 적어 자원이 제한된 환경에서도 실행이 용이합니다. 반면, B4 이상의 고버전 EfficientNet 은 자원 소모가 많아 실질적인 적용이 어려웠기 때문에, B0 가 가장 적합하다고 판단되었습니다. 더불어 EfficientNet-b0 는 전이 학습을 활용하여 적은 데이터셋에서도 높은 성능을 발휘할 수 있는 장점을 가지고 있어 실질적으로 가장 적합한 모델로 평가되었습니다.

결론적으로, EfficientNet-b0 는 태양광 발전소와 같은 산업 환경에서 효율적인 적용이 가능하며, 결함 탐지 시스템의 핵심 구성 요소로 활용하기에 가장 적합한 모델입니다.

2.3 모델 성능 개선

A 데이터 증강

데이터의 특성에 맞춰 결함의 위치가 고정되지 않았으므로 모델이 특정 방향에 치우쳐 학습되지 않고, 색상 정보의 중요도를 감소시켜 모델이 텍스처나 윤곽선선 같은 비색상적 정보를 더 잘 학습하고, 모델이 특정 특징에 의존하지 않고 전체적인 패턴을 학습하도록 유도하는 Random Horizontal Flip, Random Greyscale, Random Erasing 이 적합할 것이라고 예상하였습니다.

이후 이를 실험적인 결과와 함께 비교하여 최종 증강기법을 선택하였습니다. 실험은 배치 순서 또는 배치 구성이 바뀔때 따라 성능이 차이가 나고 그 편차가 큰 점을 고려해 각케이스에 대해 3 번의 테스트를 진행하였으며 평균과 표준편차를 이용해 비교하였습니다. 증강확률은 데이터 수가 적은점, 증강 확률이 지나치게 높으면 모델이 과적합될 가능성을 고려해 적정수준인 50%로 설정했습니다. 실험 결과 좌우반전(Random Horizontal Flip)만 적용했을 때의 성능이 가장 높았지만 표준편차가 가장 낮은건 앞서 저희가 예상했던 Random Horizontal Flip, Random Greyscale, Random Erasing 의 경우였습니다. 높은 성능도 중요하지만 안정성을 고려한 모델을 만들고자했던 프로젝트의 목적에 맞춰 최종적으로 Random Horizontal Flip, Random Greyscale, Random Erasing 증강기법을 적용하였습니다.

B Fine Tuning

전이학습을 하기 위해 먼저 모든 레이어를 freeze 한 후 상위 레이어부터 학습시키는 방식으로 진행하였습니다. 모든 레이어를 freeze 하여 pre-trained 된 모델의 가중치가 학습 도중 업데이트 되지 않게 만든 후 실험을 진행한 결과 valid loss 는

```
Test Loss: 0.4602
Test Accuracy: 85.45%
```

0.46, accuracy 는 85.45%였고 fully connected layer 이전 층들만 freeze 하고 fully connected layer 의 가중치는 학습 중 업데이트 되게 만든 경우

```
Test Loss: 1.5023
Test Accuracy: 56.97%
```

valid loss 는 1.53, accuracy 는 53.6%으로 fine tuning 을 하기 전보다 성능이 크게 떨어진 것을 확인할 수 있었습니다. 이는 학습할 수 있는 데이터양이 540 장으로 매우 적어 오히려 fine tuning 을 하면 모델이 학습을 제대로 하지 못한다고 판단하여 pre-trained 된 모델의 가중치를 모두 이용하기로 했습니다.

C Drop out

Drop out 적용 전 후 결과를 비교해본 결과 정확도와 loss 를 비교했을때 drop out 을 적용하지 않았을 때의 모델 성능이 더 좋음을 확인할 수 있습니다. 그 이유로 데이터셋의 크기가 작아 모델이 충분히 학습하기 어려운 환경이라고 판단하였습니다. Drop out 은 과적합을 방지하기 위해 사용하는 정규화 기법으로 데이터가 충분히 크지 않거나 모델이 과적합되지 않은 경우 오히려 모델의 학습 능력을 제한함을 확인할 수 있었습니다.

D Class 불균형

클래스불균형을 위해 oversampling 을 우선 적용해 보았습니다. 해당 경우 train 및 valid 에서 성능은 매우 좋게 나왔지만 test 에서의 성능은 불균형을 해소하지 않았을때와 비슷하거나 혹은 그보다 낮음을 확인했습니다. 오버샘플링은 소수 클래스의 샘플을 복제하여 데이터셋에 추가하는 방식을 이용하므로 모델이 다양한 데이터를 학습하는 것이 아닌 똑같은 데이터를 여러번 학습하게 됩니다. 따라서 모델은 훈련 데이터에 대한 적합도가 높아지지만 테스트 데이터에 대한 일반화 성능은 떨어지게 됩니다. 따라서 저희는 weightedRandomSampler 를 이용해 배치 내 클래스가 불균형하게 샘플링되는것을 방지하였습니다. WeightedRandomSampler 는 각 데이터 샘플에 가중치를 부여해 이를 기반으로 샘플링을 합니다. 즉, 소수 클래스에 더 높은 가중치를 부여해 그 클래스의 샘플이 더 많이 뽑히도록 유도합니다. 따라서 오버샘플링과 달리 중복 샘플로 인한 데이터 다양성 부족 문제를 해결하고 과적합 위험을 상대적으로 줄일 수 있었습니다.

● 배치 내 클래스 불균형					● 배치 내 클래스 불균형 해결 후				
				Std. dev Acc.					Std. dev Acc.
				2.845					2.20
Batch Size	Lr	Mean Acc.	Std. dev Acc.	Mean Loss	Batch Size	Lr	Mean Acc.	Std. dev Acc.	Mean Loss
16	0.0002	84.64	2.14	0.52	16	0.0002	83.33	1.35	0.57
	0.0005	79.42	1.43	0.66		0.0005	81.81	1.30	0.57
	0.001	77.91	5.57	0.73		0.001	78.51	2.79	0.63
32	0.0002	84.51	1.02	0.53	32	0.0002	82.38	1.61	0.62
	0.0005	82.83	2.42	0.54		0.0005	82.99	0.83	0.56
	0.001	78.31	3.62	0.71		0.001	78.40	2.08	0.71
64	0.0002	85.38	2.02	0.54	64	0.0002	84.85	1.42	0.53
	0.0005	77.41	2.94	0.69		0.0005	79.67	2.40	0.66
	0.001	72.32	4.17	0.99		0.001	73.42	3.13	0.97
128	0.0002	78.83	3.46	0.65	128	0.0002	79.35	1.49	0.69
	0.0005	75.35	3.62	0.81		0.0005	73.70	3.50	0.83
	0.001	72.26	1.73	0.99		0.001	67.07	4.46	1.26

실제로 실험을 해본 결과 실험당 모델 성능의 편차의 평균이 2.845 에서 2.20 으로 안정된 것을 확인 할 수 있었습니다. 약 0.6%의 차이는 유의미하다고 보기는 어려우나 불균형 문제를 해결 한 후 batch size = 16 일때와 batch size = 32 일 때 표준편차가 감소해 더욱 일반화가 잘 되었음을 알 수 있습니다. 다만 batch size = 64, learning rate = 0.001 일때 batch size = 128, learning rate = 0.0005/0.001 이 세가지 경우에 대해서 표준편차가 3-4%로 다른 모델들에 비해 높게 나왔습니다.

E 하이퍼파라미터

최종적으로 데이터 증강을 적용하고, fine tuning 및 drop out 을 진행하지 않고, 배치 내 클래스 불균형을 해결한 모델을 선택한 뒤 성능을 끌어올리기 위해 batch size, learning rate, learning rate scheduler 를 조정해 모델의 성능을 끌어올리려 노력했습니다. Optimizer 는 Adam 을 이용하였고 epoch 는 30 으로도 충분히 모델이 학습을 잘 하는 것으로 판단해 epoch=30 으로 고정 한 뒤 batch size = 16,32,64,128 / learning rate = 0.0002,0.0005,0.001 그리고 learning rate scheduler 는 특정 epoch 에서 학습률을 일괄적으로 감소시키는 방식과 각 epoch 마다 지정된 학습률 비율에 맞춰 학습률을 공급하는 두가지 동적 learning rate 조정 방식을 이용해 실험을 진행해 보았습니다.

● Optimizer를 ADAM으로 설정 (Lr 계산 방식 1)

Batch Size	Lr	Mean Acc.	Std. dev Acc.	Mean Loss
16	0.0002	83.33	1.35	0.57
	0.0005	81.81	1.30	0.57
	0.001	78.51	2.79	0.63
32	0.0002	82.38	1.61	0.62
	0.0005	82.99	0.83	0.56
	0.001	78.40	2.08	0.71
64	0.0002	84.85	1.42	0.53
	0.0005	79.67	2.40	0.66
	0.001	73.42	3.13	0.97
128	0.0002	79.35	1.49	0.69
	0.0005	73.70	3.50	0.83
	0.001	67.07	4.46	1.26

● Optimizer를 ADAM으로 설정 (Lr 계산 방식 2)

Batch Size	Lr	Mean Acc.	Std. dev Acc.	Mean Loss
16	0.0002	82.76	1.63	0.56
	0.0005	82.86	2.45	0.53
	0.001	80.55	2.55	0.61
32	0.0002	84.09	2.58	0.54
	0.0005	81.64	1.86	0.56
	0.001	80.03	1.79	0.67
64	0.0002	81.44	1.49	0.60
	0.0005	81.86	1.44	0.65
	0.001	78.41	3.16	0.81
128	0.0002	82.39	1.86	0.58
	0.0005	79.14	2.97	0.70
	0.001	71.98	7.81	1.04

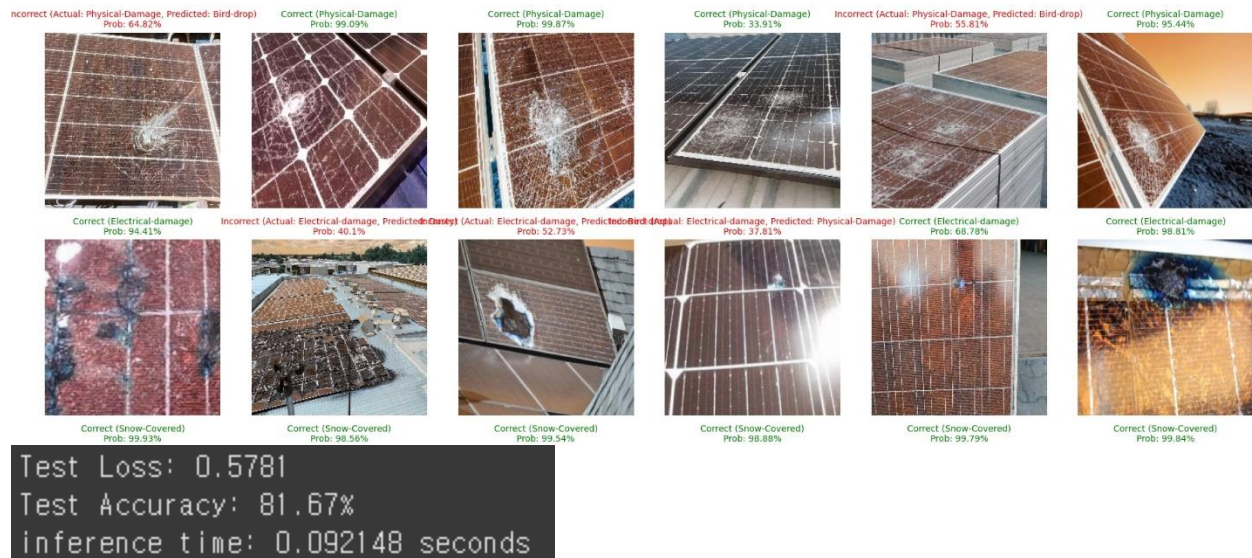
총 3 번의 실험 결과를 통해 평균과 표준편차를 구하였고, 평균과 표준편차를 바탕으로 저희 조는 단순히 수치적인 성능과 안정성뿐만 아니라 비즈니스적 관점에서 생각해보았습니다. 정확도와 안정성의 두 가지 척도로 비교를 해봤습니다. 정확도를 우선시켰을 시 비교적 일관된 데이터 환경, 즉 모델이 패널이 고장날 수 있는 대부분의 경우를 미리 학습해 놓을 수 있는 상황에서는 전체적인 효율을 극대화할 수 있겠습니다. 하지만, 태양광 패널 결함 분류에서는 신뢰도가 높아야 하며, 모델이 다양한 데이터에 대해 일관되게 동작하는 것이 중요하기에 안정성 역시 무시할 수 없었습니다. 정확도가 높더라도 특정 상황에서 성능이 크게 떨어지는 모델은 오탐/미탐으로 인해 비용 손실이나 안전 문제를 초래할 수 있습니다. 현실적으로 다른 지역, 날씨 조건, 패널 제조사에 따라 데이터가 달라지기 때문에, 다양한 환경에서 데이터가 들어올 가능성이 높습니다. 안정적인 성능은 유지보수 비용 절감에 기여하기에, 최종적으로 저희는 평균 정확도와 함께 표준편차를 함께 고려하였습니다.

이러한 과정 끝에 batch size = 64, learning rate = 0.0002, Step LR Scheduling 방식을 이용한 모델을 최종 채택했습니다.

3 결론

3.1 결과 해석 및 시각화

저희는 실제 저희가 구상한 아이디어대로 구글에서 cctv 를 통해 수집한 데이터로 추정되는 이미지들을 각 클래스 별로 10 장씩 총 60 장을 수집해 최종 모델의 성능을 평가해 보았습니다. 결과는 아래와 같았습니다.



정상 패널 데이터에 대한 분류 결과를 출력했을 때, 실제 패널에 이물질이 있는 경우인 불량을 정상으로 판단하는 경우는 적었습니다. 그 다음으로 불량으로 오분류된 정상 패널 사진을 확인했을 때, 눈을 동물의 배설물로 오분류, 카메라의 빛번짐 등으로 인한 오분류가 발생한 것을 알 수 있었습니다. 테스트 결과, 모델의 정확도는 81.67%로 나타났으며, 추론 시간은 약 0.1 초로 빠르고 효율적인 판단이 가능함을 보여주었습니다. 오분류를 줄이기 위해서는 보다 정밀한 촬영 기술의 도입과 날씨 데이터 활용을 추가적으로 고려한다면 성능을 더욱 개선할 수 있을 것으로 보입니다.

3.2 한계점

모든 연구가 완벽하게 진행되기는 다소 어려운 것처럼, 본 프로젝트 역시 진행 과정에서 몇 가지 한계점이 확인되었습니다.

가장 먼저 데이터의 제약이 본 프로젝트의 주요한 한계로 작용했습니다. 본 프로젝트는 한정된 양의 공개 데이터셋에 의존하여 실험을 진행하였는데, 이는 곧 데이터의 다양성과 양이 충분히 확보되지 않았음을 의미합니다. 이로 인해 모델의 성능을 획기적으로 향상시키는 데 한계가 있었으며, 특히 복잡한 상황에 대한 데이터가 부족하여 실제 현장에서 발생할 수 있는 다양한

상황을 충분히 반영하기 어려웠습니다. 데이터 증강 기법을 통해 가용 데이터를 다양화하려는 노력을 기울였지만, 데이터의 절대적인 양적 부족은 여전히 해결되지 않았습니다. 데이터 증강은 일부 성능 향상에 기여했지만, 기본 데이터의 품질과 양이 충분히 확보되지 않은 상태에서는 모델의 안정성과 신뢰성을 확보하는 데 한계가 있었습니다.

두 번째로 확인된 한계는 분류 기준의 모호성에서 비롯된 문제였습니다. 학습과 검증 과정에서 사용된 데이터 중 일부는 결함의 유형이 육안으로도 명확히 구분되기 어려운 경우가 포함되어 있었습니다. 이러한 애매한 데이터는 모델 또한 정확한 판단을 내리는 데 어려움을 겪었으며 분류의 신뢰성이 저하되는 모습을 보였습니다. 또한 동일한 유형의 결함이라 할지라도 결함의 정도나 형태에 따라 큰 차이가 발생하는 사례도 확인되었습니다. 이러한 데이터의 변동성은 모델이 일관된 기준으로 분류하지 못하는 원인이 되었으며, 이로 인해 모델의 정확성과 안정성이 전반적으로 저하되는 결과를 초래했습니다.

마지막으로 확인된 한계는 여러 종류의 결함이 동시에 발생할 가능성을 충분히 고려하지 못했다는 점입니다. 본 프로젝트에서 사용된 데이터는 하나의 이미지에 하나의 결함만 존재하는 경우를 전제로 했지만, 실제 현장에서는 여러 유형의 결함이 동시다발적으로 발생하는 가능성도 상당히 높을 것으로 예상됩니다. 예컨대, 물리적 손상과 먼지 오염이 함께 나타나거나, 전기적 결함과 눈 덮임이 동시에 발생하는 사례처럼 복합적인 상황이 실제로 발생할 가능성이 높습니다. 그러나 이러한 현실적인 상황을 데이터셋 구성과 모델 설계에 충분히 반영하지 못했으며, 이는 모델의 실질적인 적용 가능성과 성능을 제한하는 요인으로 작용할 수 있다는 점에서 추가적인 개선이 필요하다고 판단되었습니다.

이러한 한계점들은 연구 과정에서 불가피하게 발생하는 부분이지만, 이를 해결하고 보완한다면 본 프로젝트의 결과물이 더욱 높은 수준의 기술적 기여를 제공할 수 있을 것으로 예상됩니다.

3.3 개선 방안

앞서 확인된 한계점들은 본 프로젝트의 성과를 제한할 수 있는 중요한 요소로, 이를 해결하지 않는다면 모델의 실질적인 활용성과 효율성을 확보하기 어려울 수 있습니다. 따라서 이러한 한계점을 극복하고 모델의 성능과 적용 가능성을 한층 더 향상시키기 위해 다음과 같은 구체적인 개선 방안을 모색해 보았습니다.

우선, 다른 무엇보다도 충분한 양의 고품질 데이터 확보가 필요합니다. 선행 연구와 더불어 본 프로젝트 실험에서도 확인할 수 있듯이 데이터의 양과 질은 딥러닝 모델 성능에 직접적인 영향을 미치는 핵심 요소입니다.²⁰ 특히, 다양한 결함 유형과 실제 환경을 반영한 양질의 데이터셋을 대규모로 구축하는 것은 모델의 학습 기반을 강화하는 데 필수적입니다. 이를 통해 모델이 더 많은 상황을 학습하고 일반화 능력을 향상시킬 수 있을 뿐만 아니라, 정확성과 안정성을 포함한 전반적인 성능을 더욱 효과적으로 개선할 수 있을 것으로 기대됩니다.

다음으로는 객체 탐지 알고리즘과 하이브리드 모델의 도입을 고려해야 합니다. 객체 탐지 기술은 이미지 내 특정 영역을 강조하거나 결함이 발생한 위치를 명확히 식별하는 데 유용하며, 이를 기존의 분류 모델과 결합하면 더 높은 성능을 발휘하는 하이브리드 모델을 개발할 수 있습니다.²¹ 또한, 과거 결함 기록 데이터를 활용하면 모델이 학습 데이터의 다양성을 확보하고 판단 능력을 강화할 수 있을 것으로 예상됩니다.²² 이와 더불어, 모델이 판단을 내릴 때 확신 정도를 확률값으로 계산하여 임계치를 설정하고, 확신 정도가 임계치보다 낮은 경우에는 드론을 활용해 의심 구역을 정밀하게 촬영하거나, 고해상도 망원렌즈가 장착된 CCTV를 이용해 해당 영역을 확대하여 살펴보는 등의 기술을 결합한다면, 불확실한 판단에 대한 분류 정확도를 크게 향상시킬 수 있을 것으로 예상됩니다.²³

마지막으로 결함 분류 체계를 더욱 세분화하는 작업이 필요합니다. 결함 간 상관관계와 동시에 발생 빈도를 종합적으로 분석하여 동시 발생 가능성을 고려하고, 보다 정교한 분류 체계를 설계한다면, 동일한 유형의 결함이라고 하더라도 그 정도와 형태의 차이를 효과적으로 반영할 수 있을 것입니다. 이는 결함을 더욱 정밀하게 구분할 수 있는 기반을 제공하며, 분류 경계에서 판단이 모호한 데이터에 대해서도 높은 정확도를 유지하는 데 기여할 수 있습니다. 궁극적으로, 이러한 접근은 모델의 전반적인 신뢰성과 실질적인 적용 가능성을 한층 더 향상시키는 결과를 가져올 것으로 기대됩니다.

이와 같은 개선 방안을 통해 본 프로젝트에서 도출된 한계점을 효과적으로 극복하고 모델의 전반적인 성능과 효율성을 극대화할 수 있을 것으로 예상됩니다.

3.4 기대 효과 및 프로젝트 의의

지속 가능한 에너지 공급을 위한 핵심 기술로 자리 잡은 태양광 에너지는 여전히 발전 효율의 한계와 결함 탐지의 비효율성이라는 두 가지 주요 과제를 안고 있습니다. 본 프로젝트는 이러한 문제를 동시에 해결하기 위해 딥러닝 기술을 도입하여 기존 접근법의 한계를 극복하고, 태양광 발전 시스템의 성능과 관리 효율성을 혁신적으로 향상시키는 새로운 방식을 제안하였습니다.

우선, 본 프로젝트는 기존 재료와 공정 기술, 발전 원리 중심으로 이루어진 접근법이 가진 한계를 보완하여 발전 효율을 극대화할 가능성을 모색하였습니다. 특히, 딥러닝 기술을 활용한 새로운 방법론은 태양광 패널 표면을 정밀하게 분석하여 결함을 신속하고 정확하게 탐지하고, 결함 유형을 체계적으로 분류하는 기능을 제공합니다. 이를 통해 작은 결함이 심각한 문제로 확대되기 전에 조기 탐지가 가능하며, 선제적 대응을 통해 고장 발생 빈도를 낮추고 예상치 못한 운영 중단을 효과적으로 방지할 수 있습니다. 결과적으로, 본 모델은 태양광 발전소의 안정적이고 지속적인 운영을 지원하며, 발전 효율을 극대화하는 동시에 발전 단가 절감이라는 두 가지 목표를 달성하는 데 기여할 것입니다.

태양광 발전 시스템의 관리 측면에서도 본 프로젝트의 성과는 주목할 만합니다. 본 프로젝트에서 제시한 모델은 정확하고 효율적인 관리 시스템을 제공하여 관리 시간과 비용을 절감하고, 인적·물적 자원의 투입을 최소화하도록 설계되었습니다. 특히, 사막과 같은 접근이 어려운 외진 지역에 최근 잇따라 건설되고 있는 대규모 발전소에서 본 모델의 가치는 더욱 두드러질 것으로 예상됩니다. 기존의 관리 및 결함 탐지 방식은 대규모 발전소의 넓은 면적과 복잡한 유지·보수 요구를 효과적으로 해결하지 못해 높은 관리 비용과 비효율성을 초래했습니다. 반면, 본 프로젝트의 모델은 이러한 문제를 극복하며 발전소 운영의 효율성을 극대화할 것입니다. 나아가, 기회비용을 줄이고 자원 활용을 최적화함으로써 발전소가 높은 생산성을 실현할 수 있도록 기여할 것입니다.

이러한 딥러닝 기반의 새로운 접근법은 태양광 발전소 운영 방식의 패러다임을 전환시킬 잠재력을 가지고 있습니다. 발전 효율을 극대화하고 관리 비용을 획기적으로 절감함으로써 단순한 운영상의 개선을 넘어, 태양광 발전의 경제성과 지속 가능성을 한층 강화할 수 있습니다. 결론적으로 본 프로젝트는 태양광 발전의 신뢰성과 효율성을 높은 수준으로 끌어올릴 수 있는 계기를 마련하며, 이를 바탕으로 안정적이고 지속 가능한 발전 환경을 구축함으로써 진정한 탄소 중립 시대의 실현이라는 궁극적 목표에 한 걸음 더 가까이 다가가는 데 기여할 것으로 기대됩니다.

4 Reference

¹ Bahar, H., et al. (2024). Renewables 2024: Analysis and forecast to 2030. *International Energy Agency*.

² Çam, E., et al. (2024). Electricity 2024: Analysis and forecast to 2026. *International Energy Agency*.

³ Shockley, W., & Queisser, H. J. (1961). Detailed balance limit of efficiency of p-n junction solar cells. *Journal of Applied Physics*, 32(3), 510-519.

⁴ Polman, A., et al. (2016). Photovoltaic materials: Present efficiencies and future challenges. *Science*, 352(6283), 307–307.

⁵ Massoud, Y., et al. (2023). Fourth-generation solar cells: A review. *Energy Advances*, 2(9), 1239–1262.

⁶ Rapaport, S., et al. (2021). The use of advanced algorithms in PV failure monitoring. *IEA PVPS Task 13 Report. International Energy Agency*.

⁷ Woo, S., et al. (2022). A comparative study on the machine learning model for anomaly detection in solar photovoltaics systems. *Proceedings of the 2022 Korea Software Congress*, 690-692.

⁸ Baltacı, Ö., et al. (2024). Thermal image and inverter data analysis for fault detection and diagnosis of PV systems. *Applied Sciences*, 14(9), 3671.

⁹ Köntges, M., et al. (2017). Assessment of photovoltaic module failures in the field. *IEA PVPS Task 13 Report IEA-PVPS T13-09:2017. International Energy Agency*.

¹⁰ Wainwright, O. (2023, May 21). *How solar farms took over the California desert: 'An oasis has become a dead sea'*. The Guardian. <https://www.theguardian.com/us-news/2023/may/21/solar-farms-energy-power-california-mojave-desert>

¹¹ Yang, J., & Li, J. (2017). Application of deep convolution neural network. *Proceedings of the 2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 229–232.

¹² Kingma, D. P., et al. (2013). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.

¹³ Goodfellow, I., et al. (2014). Generative Adversarial Networks. *arXiv preprint arXiv:1406.2661*.

¹⁴ Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations*.

¹⁵ Herz, M., et al. (2021). Quantification of technical risks in PV power systems. *IEA PVPS Task 13 Report IEA-PVPS T13-23:2021. International Energy Agency*.

¹⁶ Simonyan, K., et al. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

¹⁷ Howard, A., et al. (2019). MobileNetV3: An efficient architecture for mobile vision. *arXiv preprint arXiv:1905.02244*.

¹⁸ Tan, M., et al. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning, PMLR 97*.

¹⁹ Huang, G., et al. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.

²⁰ Alom, M. Z., et al. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292.

²¹ Girshick, R., et al. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 580–587.

²² Shi, X., et al. (2019). Hybrid convolutional recurrent neural networks outperform CNN and RNN in task-state EEG detection for Parkinson's disease. *Proceedings of APSIPA Annual Summit and Conference 2019*, 939–944.

²³ Corbière, C., et al. (2019). Addressing failure prediction by learning model confidence. *NeurIPS 2019 Conference Proceedings*.