

# CNN 을 이용한 Multi-Scale 웨이퍼 결함 분류

이상현, 김동환, 조용진, 고현석\*  
한양대학교 ERICA 전자공학부

{roylee2001, donghk99, choyj324, hyunsuk}@hanyang.ac.kr

## Multi-Scale Wafer Defect Classification Using Convolutional Neural Networks

Sang-Hyun Lee, Dong-Hwan Kim, Yong-Jin Jo, Hyunsuk Ko  
School of Electrical Engineering, Hanyang University ERICA

### 요약

본 연구에서는 반도체 테스트 공정 시에 발생하는 데이터인 WBM(Wafer Bin Map)이 불균형하고 다중 크기를 가진다는 특성을 고려하여, 세 가지 이미지 전처리 기법—(1) 리사이즈(Resize), (2) 3-채널 원-핫 인코딩(Pixel-wise One-Hot Encoding), (3) 오토인코더 기반 데이터 증강(Auto-encoder Based augmentation)—을 CNN(Convolution Neural Network) 학습 전에 적용하였다. 이후 배치 정규화(Batch Normalization)와 드롭아웃(Dropout)을 포함한 경량 CNN 을 설계하였다. 이는 전처리 된 이미지의 학습에 적용되어 효과적으로 과적합을 억제하고 94.45%의 테스트 정확도를 달성하였다.

### 1. 서론

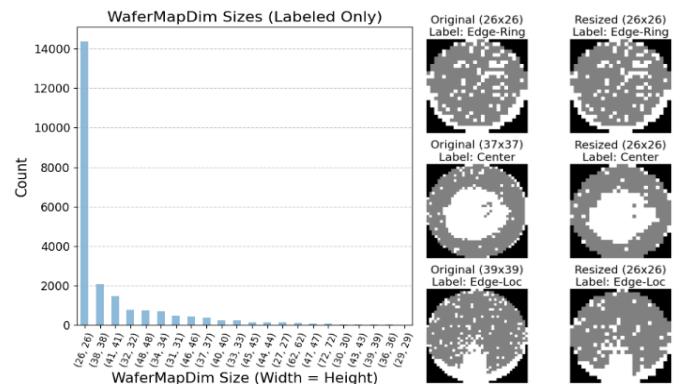
반도체 제조 공정에서는 EDS(Electrical Die Sorting) 테스트를 통해 웨이퍼 칩의 전기적 특성을 검증하고, 결과를 WBM(Wafer Bin Map) 형태로 기록한다.[1] 기록된 WBM의 분석은 수율 향상을 위한 중요한 절차이다.[2] 그러나, WBM의 불량 유형(Label)은 대부분의 경우 불균일하다. 이에 본 논문은 이미지 전처리 단계에서 세 가지 방법을 적용하여, 다중 크기의 웨이퍼 결함을 CNN(Convolution Neural network)로 분류하였다. 적용된 전처리 기법은 (1)  $26 \times 26$  크기로 WBM 리사이즈(Resize) (2) 3채널 원-핫 인코딩 (One-hot Encoding) (3) 오토인코더 기반 데이터 증강(Auto-encoder Based augmentation)이다.

### 2. 본론

#### 2.1 WBM 입력 데이터 리사이즈

본 연구에서는 Kaggle의 WM-811k 데이터셋을 사용하였다. WM-811k은 46,393개 공정에서 811,457개의 WBM을 추출했으며, 9개의 불량 유형(Label: Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-full, None)으로 분류된다. 불량 유형이 존재하는

WBM 크기를 조사한 결과,  $26 \times 26$ 이 약 14,000개로 가장 많았다(그림 1). 이 크기만을 이용해 불량 유형을 분류한다면, 데이터 다양성이 제한되고 모델 과적합 문제를 초래한다. 이를 해결하기 위해 본 논문은 OpenCV library의 최근접 보간법을 통해 모든 정사각형 WBM을  $26 \times 26$  크기로 리사이즈 하였다(그림 1)

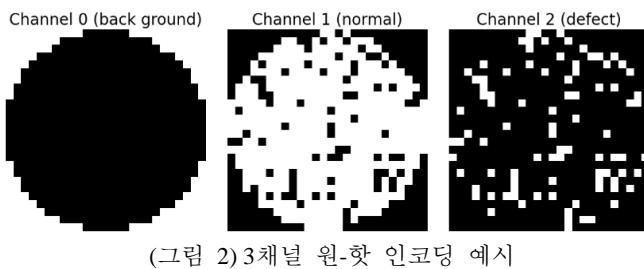


(그림 1) WBM 다중 크기 분포 시각화 및 리사이즈 예시

#### 2.2.3 채널 확장을 위한 원-핫 인코딩

기존 WBM 데이터는 픽셀 하나에 0(공백), 1(정상), 2(불량) 값을 모두 포함하는 단일 채널 형태로 저장되

어 있었다. 이전 연구는 불량 유형에 대해서만 원-핫 인코딩을 진행하였고, 이미지 픽셀 값에는 진행하지 않았다.[3] 본 연구에서는 이미지 전처리 단계에서 0(공백), 1(정상), 2(불량)을 각각 별도의 채널로 분리하는 3채널 원-핫 인코딩을 적용하였다(그림 2). 이는, WBM의 픽셀 값은 연속적인 값이 아닌 범주형 값이라는 것에 기반한다. 결론적으로, 불필요한 우선 순위를 없애고 각 불량 유형을 명확히 구분할 수 있었다.



(그림 2) 3채널 원-핫 인코딩 예시

### 2.3 오토인코더 기반 데이터 증강

데이터 증강 필요성을 확인하기 위해 9개 불량 유형에 대해 WBM 수를 분석한 결과, None은 19,191개로 압도적으로 많았고, Donut과 Near-full은 각각 14개, 32개로 매우 적었다(그림 3). 이러한 데이터 불균형 문제를 해결하기 위해 본 연구는 오토인코더 기반 증강 기법을 적용하였다. 원-핫 인코딩된 WBM을 인코더(Encoder)를 통해 잠재 공간(Latent Space)으로 압축하고 가우시안 노이즈(Gaussian noise)를 추가한 뒤, 디코더(Decoder)로 복원하여 데이터를 균일하게 확장하였다(그림 3). 본 방법은 회전, 반전 등 여러 기법의 혼합 없이 하나의 절차로 간편히 수행되며, 이미지의 변형 없이 증강이 가능하다.

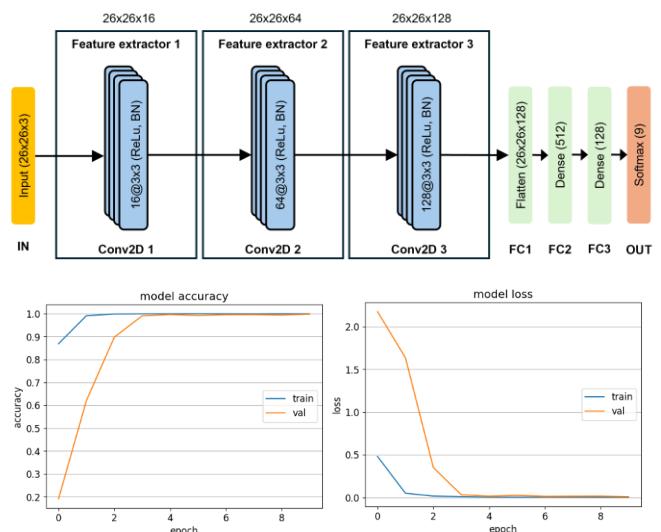
Label	# of wafer bin map	Ratio [%]	After augmentation	Ratio after augmentation [%]
Center	803	3.51%	5,778	11.69%
Donut	14	0.06%	5,016	10.15%
Edge-Loc	680	2.97%	5,984	12.10%
Edge-Ring	1,093	4.77%	6,118	12.37%
Loc	634	2.77%	5,577	11.28%
Near-full	32	0.14%	5,044	10.20%
Random	282	1.23%	5,424	10.97%
Scratch	165	0.72%	5,148	10.41%
none	19,191	83.83%	5,353	10.83%
Total	22,894		49,442	

(그림 3) 오토인코더 기반 데이터 증강 전후의 WBM 불량 유형 분포

### 3. 결론

본 논문에서 제안하는 CNN 모델의 구조는 3차원 WBM ( $26 \times 26 \times 3$ )을 입력으로 하며, 3개의 Convolution 층, 1개의 Flatten 층, 2개의 Dense 층으로 구성된다(그림 4). Convolution 층은  $3 \times 3$  kernel 필터를 16, 64, 128개를 사용하였으며, 배치 정규화(Batch normalization)를

통해 입력 분포의 균일성을 확보하였다.[3] Dense 층은 각각 512, 128개의 노드로 구성되었으며, 30% 드롭아웃(Dropout)을 통해 노드를 제한하여 과적합을 방지했다.[3] 활성화 함수는 모든 층에서 동일하게 RELU를 사용하였다. 데이터의 70%를 학습(Train), 20%를 검증(Validation), 10%를 시험(Test) 용도로 분리하여 성능 평가를 진행했다. 20 epoch 학습 결과, 학습용과 검증용 데이터의 정확도는 각각 99.3%, 98.9%, 손실은 0.0235, 0.0391을 기록하였다. 시험용 데이터에서는 정확도 94.45%, 손실 0.2589가 도출되었다. 이는 제안한 WBM 전처리를 통해 CNN 모델이 적은 데이터에서도 과적합 없이 높은 분류 정확도를 달성할 수 있음을 시사한다.



(그림 4) 제안한 CNN 모델 구조 및 학습/검증 정확도와 손실 변화

### 참고문헌

- [1] K. Kyeong and H. Kim, "Classification of Mixed-Type Defect Patterns in Wafer Bin Maps Using Convolutional Neural Networks," in IEEE Transactions on Semiconductor Manufacturing, vol. 31, no. 3, pp. 395-402, Aug. 2018
- [2] Seung-Hyun Choi, Dong-Hee Lee, Eun-Su Kim, Young-Mok Bae, Young-Chan Oh, Kwang-Jae Kim, Development of a spatial dimension-based taxonomy for classifying the defect patterns in a wafer bin map, Advanced Engineering Informatics, Volume 60, 2024
- [3] M. Saqlain, Q. Abbas and J. Y. Lee, "A Deep Convolutional Neural Network for Wafer Defect Identification on an Imbalanced Dataset in Semiconductor Manufacturing Processes," in IEEE Transactions on Semiconductor Manufacturing, vol. 33, no. 3, pp. 436-444, Aug. 2020