

점진적 데이터 평준화를 이용한 반도체 웨이퍼 영상 내 결함 패턴 분류

도정혁, 김문철

한국과학기술원 전기 및 전자 공학부

ehwjdgur0913@kaist.ac.kr, mkimee@kaist.ac.kr

Wafer Map Defect Pattern Classification with Progressive Pseudo-Labeling Balancing

Jeonghyeok Do, Munchurl Kim

Korea Advanced Institute of Science and Technology Dep. of Electronic Engineering

요 약

전 반도체 제조 및 검사 공정 과정을 자동화하는 스마트 팩토리의 실현에 있어 제품 검수를 위한 검사 장비는 필수적이다. 하지만 딥 러닝 모델 학습을 위한 데이터 처리 과정에서 엔지니어가 전체 웨이퍼 영상에 대하여 결함 항목 라벨을 매칭하는 것은 현실적으로 불가능하기 때문에 소량의 라벨 (labeled) 데이터와 나머지 라벨이 없는 (unlabeled) 데이터를 적절히 활용해야 한다. 또한, 웨이퍼 영상에서 결함이 발생하는 빈도가 결함 종류별로 크게 차이가 나기 때문에 빈도가 적은 (minor) 결함은 잡음처럼 취급되어 올바른 분류가 되지 않는다. 본 논문에서는 소량의 라벨 데이터와 대량의 라벨이 없는 데이터를 동시에 활용하면서 결함 사이의 발생 빈도 불균등 문제를 해결하는 점진적 데이터 평준화 (progressive pseudo-labeling balancer)를 제안한다. 점진적 데이터 평준화를 이용해 분류 네트워크를 학습시키는 경우, 기존의 테스트 정확도인 71.19%에서 6.07%-p 상승한 77.26%로 약 40%의 라벨 데이터가 추가된 것과 같은 성능을 보였다.

1. 서론

혁신적인 반도체 검사 장비의 한 축은 머신 비전을 바탕으로 하는 딥 러닝 기반 반도체 웨이퍼 결함 분류 기술이다. 결함 분류 알고리즘의 성능 향상으로 훨씬 효율적이고 고신뢰성의 반도체 생산 공정 구축에 크게 기여할 수 있다. 웨이퍼 결함 분류의 경우, 딥 러닝 모델 학습을 위한 데이터 처리 과정에서 엔지니어가 전체 웨이퍼 영상에 대하여 결함 항목 라벨을 매칭하는 것은 현실적으로 불가능하기 때문에 소량의 라벨 데이터와 나머지 라벨이 없는 데이터를 효율적으로 활용하는 것이 필수적이다. 또한, 웨이퍼 영상에서 결함이 발생하는 빈도가 결함 종류별로 크게 차이가 나기 때문에 네트워크 학습 시 발생 빈도가 적은

결함은 잡음처럼 취급되어 올바른 분류가 되지 않는다.

웨이퍼 결함 패턴 분류에 관한 선행 연구로, Nakazawa [1]는 푸아송 점 과정 (Poisson point process)을 통해 웨이퍼 결함 패턴을 인위적으로 생성하여 라벨 데이터 부족 문제를 해결하는 시도를 하였으며, Wang [2]은 결함 클래스 간의 데이터 개수 불균등으로 원활하게 학습이 되지 않는 문제를 해결하기 위해 GAN 을 도입하여 결과 정확도가 상대적으로 낮은 클래스에 해당하는 데이터를 더 많이 생성하는 알고리즘을 제안한다. 또한, Yang [3]은 소량의 라벨 데이터와 대량의 라벨이 없는 데이터로부터 불량 패턴을 분류하는 네트워크를 학습하기 위하여 준지도학습 (semi-supervised learning)을 도입하여 라벨 데이터셋의 크기를 키움으로써 네트워크의 오버피팅 (overfitting)을 방지하였다.

다양한 연구에도 불구하고 (1) 라벨 데이터의 부족, (2) 결함 패턴 클래스 간의 데이터 개수 불균등, 네트워크 학습 시 두 가지의 문제를 동시에 해결하기 위한 연구는 아직 미비하다. 본 논문에서는 풍부한 라벨이 없는 데이터를 활용하여 위 두 가지의 문제를 모두 해결하는 새로운 학습 전략인 점진적 데이터 평준화를 제안한다. 소량의 라벨 데이터로 효과적인 네트워크 학습법을 도출해낸다면 엔지니어의 라벨링 (labeling) 업무 효율화에 크게 기여할 것으로 예상된다.

2. 불균등 오버-샘플러 (Imbalanced over-sampler)

점진적 데이터 평준화가 적용되기 전, 근본적인 결함 라벨 간의 불균등 문제를 완화하기 위해 데이터가 적은 클래스에 데이터를 추가해주는 오버-샘플링 기법을 변형하여 적용하였다. 오버-샘플링과 동일한 효과를 얻기 위해, 네트워크 학습에서 배치 (batch)를 얻을 시, 하나의 배치 내부에 포함된 클래스별 데이터가 균등하도록 샘플링하는 배치 샘플러 (batch sampler)인 불균등 오버-샘플러를 설계하였다.

3. 점진적 데이터 평준화 (Progressive pseudo-labeling balancer)

본 연구에서 제안하는 점진적 데이터 평준화는 라벨이 없는 데이터를 라벨링 하여 라벨 데이터로 사용하는 가(假) 정답 제작 (pseudo-labeling) 방법을 기반으로 라벨이 없는 데이터를 유의미하게 활용하여 결함 라벨 간의 불균형 문제를 해결한다.

3.1. 초기 학습 단계 (True label based training phase)

Figure. 1 에서 소량의 라벨 데이터와 불균등 오버-샘플러로 결함 분류 네트워크를 학습한다. 학습된 네트워크가 가(假) 정답을 제작해야 하므로 분류 성능을 최대한 높여 올바른 정답이 예측할 수 있도록 학습되어야 하며, 이를 위해 기존의 불균등 학습법인 focal loss [4]나 그 외의 방법 [5, 6]을 도입할 수 있다.

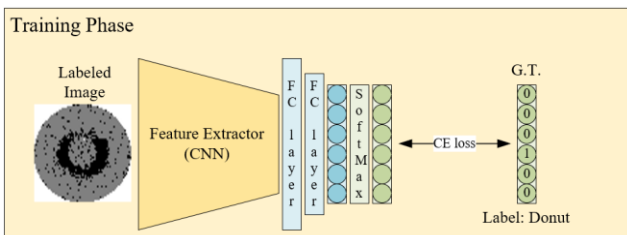


Figure. 1. 초기 학습 단계

3.2. 가(假) 정답 제작 단계 (Pseudo-labeling phase for training data augmentation)

Figure. 2 에서 학습된 네트워크에 라벨이 없는 데이터를 입력으로 가(假) 정답 제작하여 학습을 위한 라벨 데이터를 증대한다. 네트워크 출력은 소프트맥스 (Softmax) 함수를 통과하였으므로 모든 원소의 합이 1 인 확률 벡터로 나타난다. 확률벡터의 원소 중 가장 큰 값을 갖는 위치를 가(假) 정답으로 설정한다. 같은 가(假) 정답을 가지더라도 그 라벨이 얼마큼 정확하지에 대한 신뢰도는 모두 다르게 된다. 이 차이를 고려해 주기 위해 소프트맥스 함수를 통과한 최종 확률 벡터의 엔트로피 (entropy) [7]를 계산 및 저장하여 이후 네트워크 학습 시에 손실함수 (loss function)의 가중치로 사용하였다.

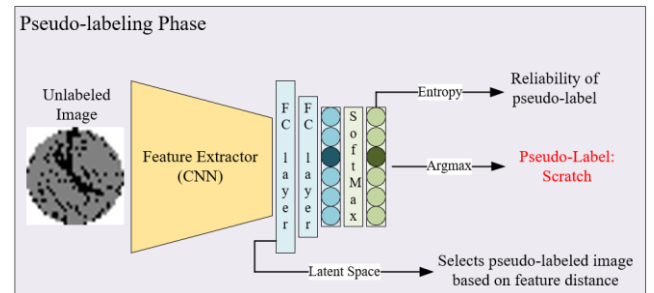


Figure. 2. 가(假) 정답 제작 단계

3.3. 평준화 단계 (Balancing phase for training data augmentation)

기존 데이터셋에 가(假) 정답 데이터를 선택적으로 추가하여 데이터 클래스 간의 균형을 맞추는 평준화 단계를 진행한다. 잠재 공간 (latent space) 상에서 특징 지도 (feature map)의 종류가 다양해지면 네트워크가 더욱 여러 가지 특징 지도에 대해 견고하게 학습이 된다. 따라서, 라벨 데이터 중에서 네트워크가 올바르게 판단한 데이터의 출력 확률 벡터 최대값의 분포는 유지하면서 특징 지도의 분포를 풍부하게 하는 가(假) 정답 데이터를 선택하였다. 예를 들어, 특정 클래스에 대해 네트워크가 맞게 분류한 데이터가 3 개이고, 출력 확률 벡터의 최대값이 각각 0.9, 0.9, 0.8 이라면, 출력 확률 벡터 최대값이 0.9, 0.8 과 가까운 가(假) 정답 데이터를 2:1 비율로 선택하여 기존 데이터셋에 추가한다. 이때, 비슷한 출력 확률 벡터 최대값을 갖는 가(假) 정답 데이터가 여러 개 존재한다면, 특징 지도 간의 거리를 측정하여 특징 지도의 분포를 풍부하게 하는 데이터를 선택한다. 잠재 공간상에서 특징 지도 사이의 거리로 프로베니우스 노름 (Frobenius norm)을 사용하였다.

클래스별 최종 데이터 수, Max_num 을 설정하고, 매 평준화 단계마다 일정한 수의 데이터를 Figure. 3 과 같이 각 클래스에 추가해준다. 따라서 평준화 단계가 반복될수록 데이터 분포가 점차 균등해지게 된다.

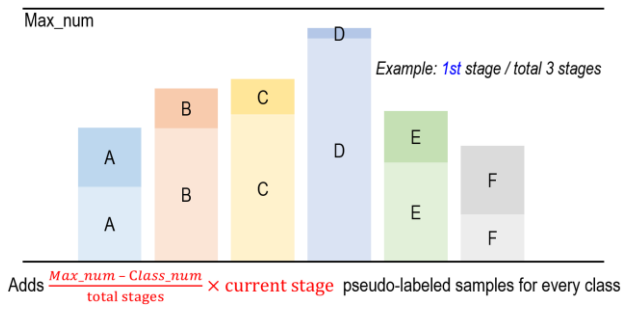


Figure 3. 평준화 단계 (현재 1 단계 / 전체 3 단계)

4. 실험 결과

4.1. 실험 환경 설정

모든 실험에서 아담 최적화 기법 (Adam optimizer)을 사용하였고, 초기의 학습률 (learning rate)는 $1e-3$ 로 설정한 뒤, 50 에폭 (epoch)마다 절반으로 감소시켰다. 배치 크기는 128 로 설정하였다. 근본적으로 발생하는 데이터 분포의 불균형 문제를 해결하기 위해 배치 샘플링 시 불균등 오버-샘플러를 기본으로 사용하였다. 회전 (rotation)과 뒤집기 (flip)의 데이터 증대 기법을 사용하였고, 입력 웨이퍼 영상의 크기는 가로, 세로를 동일하게 32 로 재설정하여 사용하였다. 손실 함수로는 교차 엔트로피 (cross entropy) 손실 함수를 사용하였다.

처음 150 에폭은 라벨 데이터만으로 학습을 진행하는 초기 학습 단계, 다음 150 에폭은 가(假) 정답 제작 단계, 그리고 평준화 단계를 통해 얻은 가(假) 정답 데이터와 라벨 데이터 모두를 포함하는 학습 데이터를 이용하여 학습하였다. 마지막 50 에폭은 최종 가(假) 정답 데이터와 라벨 데이터, 즉 최대 규모의 데이터를 이용하여 학습을 안정적으로 수행하였다.

Tarvainen [8]의 13-계층 콘볼루션 네트워크 (13-layer ConvNet)를 기반으로 실험을 진행하였다. 150 (=150+0), 200 (=150+50), 250 (=150+100) 에폭에서 가(假) 정답 제작 단계, 평준화 단계를 총 3 번 진행하였다. 클래스별 최종 데이터 수, Max_num 은 100 으로 하였고, 라벨이 없는 데이터의 개수, 특징 지도의 거리를 비교할 대상은 다양하게 변화시켜가며 실험을 진행하였다.

4.2. 데이터셋 구성

실제 반도체 제조과정에서 취득한 반도체 웨이퍼 영상 811,457 장으로 구성된 WM-811K 데이터셋을 사용하여 실험을 진행하였다. 나타나는 결함 패턴의 종류는 총 9 가지로 Figure. 4 [9]와 같다.

불균등 비율 (imbalance ratio)이 100 이 되도록 인위적으로 WM-811K 에서 데이터셋을 구축하였다 [6]. 가장 작은 클래스는

'Near-full'로 10 장의 표본을 갖고, 가장 큰 클래스는 'None'으로 1,000 장의 샘플을 포함한다. 학습에 총 2,275 장의 라벨 데이터를 사용하였으며, 평가에는 모든 클래스의 표본 수가 95 장으로 총 855 장을 사용하였다. 평가 데이터가 균등하기 때문에 전체 (평균) 정확도 (accuracy)로 성능을 평가하였다.

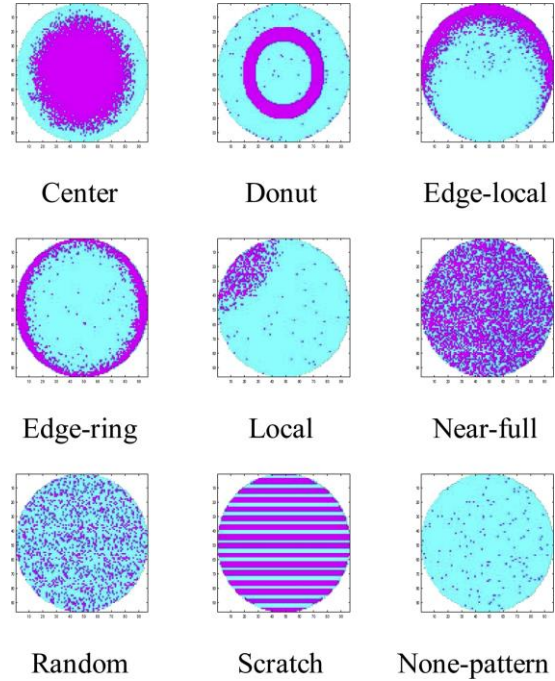


Figure 4. WM-811K 의 결함 패턴

4.3. 점진적 데이터 평준화 적용 결과

라벨이 없는 데이터, UL (=UnLabeled) 를 {10,000, 20,000, 30,000, 40,000}, 특징 지도의 거리를 비교할 대상, NB (=NeighBour)을 {5, 10}으로 설정하고 실험을 진행하였다. 공평한 비교를 위해 점진적 평준화를 사용하지 않고 350 (=150+150+50) 에폭만큼 일반적인 지도학습을 적용한 것을 비교 대상으로 하였다.

NB \ UL	지도학습	10,000	20,000	30,000	40,000
5		77.26%	73.81%	76.55%	73.21%
	71.19%	(200)	(166)	(278)	(274)
10	(129)	73.57%	75.48%	75.48%	73.57%
		(165)	(243)	(225)	(174)

Table 1. 점진적 데이터 평준화 적용 결과 (최고 정확도와 해당 에폭)

Table. 1 에서 점진적 데이터 평준화를 적용하였을 때, 크기는 6%-p 이상의 정확도 향상을 보였다. 지도학습만으로 네트워크를 학습하는 경우, 라벨 데이터의 부족으로 학습 데이터의 분류 정확도는 100%에 가깝지만 평가 데이터의 분류 정확도는 학습 초기에 최고점에 도달하고 그 이후에 하락하는

오버피팅 문제가 발생하였음을 알 수 있다. 점진적 데이터 평준화를 적용하여 가(假) 정답 데이터를 계속 제공하면 데이터 증대로 이러한 오버피팅 문제를 완화할 수 있기 때문에 성능 개선에 도움이 되었다. 또한, 네트워크가 학습할 수 있는 데이터가 풍부해지는 것 역시 성능 향상에 기여를 하였다.

4.4. 라벨 데이터에 따른 일반적인 지도 학습과 비교

점진적 데이터 표준화의 효과를 확인하기 위해, 라벨 데이터 수를 조절해가며 350 에폭동안 일반적인 지도학습으로 네트워크를 학습시켰다.

라벨 데이터	1,139	2,275	3,411	4,547
정확도	67.14% (133)	71.19% (129)	79.52% (94)	81.79% (127)

Table. 2. 라벨 데이터 개수에 따른 결과 (최고 정확도와 해당 에폭)

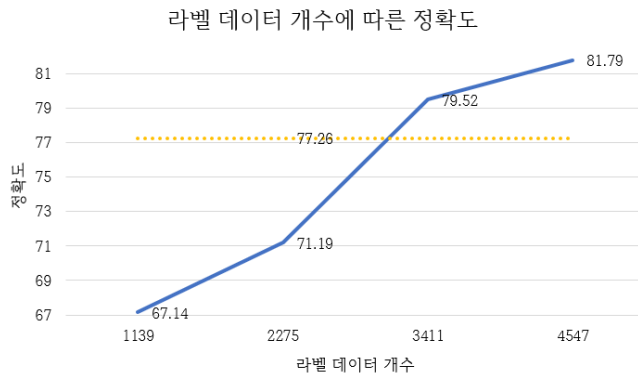


Figure. 5. 라벨 데이터 개수에 따른 정확도

Table. 2 에서 라벨 데이터가 많아질수록 성능이 확실하게 개선되는 것을 확인할 수 있다. Figure. 5 는 Table. 2 를 토대로 라벨 데이터와 정확도 사이의 관계를 도시한 것이다. 점진적 데이터 표준화를 적용하면 약 40%의 라벨 데이터가 추가된 것과 같은 성능을 보인다.

5. 결론

반도체 웨이퍼 데이터의 경우, 라벨 데이터가 부족하고 결함 패턴 클래스 간의 데이터 개수 불균등하기 때문에 네트워크 학습에 어려움이 따른다. 위의 문제점들을 극복하기 위해 데이터의 불균형 문제를 해결하고 동시에 라벨이 없는 데이터를 적절히 활용하는 새로운 학습 전략인 점진적 데이터 평준화를 제안하였다. 적용 결과, 동일한 네트워크에서 6.07%-p 의 분류 정확도가 향상되어 77.26%로 약 40%의 라벨 데이터가 추가된 것과 같은 성능을 보였다.

Acknowledgements

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00419, Intelligent High Realistic Visual Processing for Smart Broadcasting Media)

참고문헌

- [1] T. Nakazawa and D.-V. Kulkarni, "Wafer Map Defect Pattern Classification and Image Retrieval Using Convolutional Neural Network," In IEEE Transactions on Semiconductor Manufacturing, vol. 31, no. 2, pp. 309-314, May. 2018.
- [2] J. Wang, Z. Yang, J. Zhang, Q. Zhang and W.-K. Chien, "AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition," IEEE Transactions on Semiconductor Manufacturing, vol. 32, no. 3, pp. 310-319, Aug. 2019.
- [3] S.-B. Yang and T.-L. Yu, "Pseudo-representation labeling semisupervised learning," arXiv, pp. arXiv-2006, May. 2020.
- [4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," In ICCV, Oct. 2017.
- [5] N.Chawla, K. Bowyer, L. Hall and W. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal Of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
- [6] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," In CVPR, 2019.
- [7] A. Iscen, G. Tolias, Y. Avrithis, and O. Chum, "Label propagation for deep semi-supervised learning," In CVPR, 2019.
- [8] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semisupervised deep learning results," In NeurIPS, pp. 1195-1204, 2017.
- [9] J. Yu, X. Zheng, and J. Liu, "Stacked convolutional sparse denoising autoencoder for identification of defect patterns in semiconductor wafer map," Computers in Industry, vol. 109, pp. 121-133, Aug. 2019.