

신생아 호흡기 질환의 정밀 진단을 위한 영상 문맥 기반의 특징 게이팅 네트워크 연구

A Study on Clinical-Context Gated Network for Precision Diagnosis of Neonatal Respiratory Diseases

정수민¹, 최병민^{2*}

Sumin Jung¹, Byung Min Choi^{2*}

요약

신생아 중환자실(NICU)에서 흉부 X-ray(CXR)는 호흡기 질환 진단의 핵심 도구이다. 선행 연구에서는 딥러닝 모델에 재태 주수와 출생 체중 정보를 단순 결합(Early Fusion)하여 진단 성능을 개선하였으나, 이는 고차원 영상 공간에서 단일 임상 수치가 희석되는 정보 밀도의 불균형(Information Density Imbalance)과 특징 간 상호작용 부재라는 한계를 가진다. 본 연구에서는 임상 정보가 영상 특징 추출을 동적으로 제어하는 Clinical-Context Gated Network (CCGN)를 제안한다. 제안 모델은 임상 정보를 바탕으로 채널별 어텐션(Channel-wise Attention)을 생성하여 질환 판별에 중요한 영상 특징을 선별적으로 강조한다. 43,338건의 대규모 데이터셋을 이용한 실험 결과, 제안 모델은 ResNet50(Baseline, F1-score 0.8368, Macro-AUROC 0.9122)과 선행 Early Fusion 기반 모델(Previous Work, F1-score 0.8632, Macro-AUROC 0.9352) 대비 F1-score 0.8728, Macro-AUROC 0.9457의 성능을 달성하였으며, ConvNeXt-Tiny 및 Swin-T 등 최근 제안된 일반 목적 백본(backbone)과 비교해도 일관된 우위를 보였다. 특히 진단이 까다로운 신생아 일과성 빈호흡(TTN)의 분류 정확도를 크게 향상시켰다.

핵심어 : 딥러닝, 신생아 호흡기 질환, 멀티모달 융합, 어텐션 메커니즘, 의료 영상 분석

Abstract

In neonatal intensive care units (NICUs), chest radiography (CXR) is a primary tool for diagnosing respiratory diseases. Previous deep-learning studies have improved diagnostic performance by simply concatenating gestational age (GA) and birth weight (BW) with image inputs via early fusion. However, this strategy suffers from information density imbalance in the high-dimensional image space, where a few scalar clinical variables are diluted, and it does not explicitly model interactions between image features and clinical context. In this study, we propose a Clinical-Context Gated Network (CCGN) in which clinical information dynamically modulates image feature extraction. CCGN generates channel-wise attention weights

1 Core Research&Development Center, Korea University Ansan Hospital, Ansan-si, Republic of Korea [Research Professor]
e-mail: sumini1019@korea.ac.kr

2 Department of Pediatrics, Korea University Ansan Hospital, Ansan-si, Republic of Korea [Professor]
e-mail: cbmin@korea.ac.kr (Corresponding author)

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2024-00457819)

Received(November 08, 2025), Review Result(1st: November 26, 2025), Accepted(December 12, 2025), Published(December 31, 2025)



© 2025 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

from clinical variables and selectively emphasizes image features that are critical for disease discrimination. Using a large-scale multicenter dataset of 43,338 neonatal CXRs, the proposed model achieved an F1-score of 0.8728 and a macro-area under the receiver operating characteristic curve (Macro-AUROC) of 0.9457, outperforming both the ResNet50 baseline (F1-score 0.8368, Macro-AUROC 0.9122) and a previous early-fusion model (F1-score 0.8632, Macro-AUROC 0.9352). CCGN also showed consistently superior performance compared with recently proposed generic backbones such as ConvNeXt-Tiny and Swin-T. Notably, CCGN substantially improved the classification performance of transient tachypnea of the newborn (TTN), which is particularly challenging to diagnose, demonstrating its potential as a clinically useful decision-support tool in NICU practice.

Keyword : deep learning, neonatal respiratory diseases, multimodal fusion, attention mechanism, medical image analysis

1. 서론

1.1 연구 배경

신생아 호흡곤란증후군(RDS), 신생아 일과성 빈호흡(TTN), 기관지폐이형성증(BPD) 등의 호흡기 질환은 신생아 중환자실(NICU) 입원 환자의 주요 사망 원인 중 하나이다. 흉부 방사선 촬영(CXR)은 이러한 질환을 진단하는 일차적 도구이나, 질환 간의 영상학적 유사성으로 인해 숙련된 전문의 조차 감별 진단에 어려움을 겪는다 [1]. 이에 딥러닝을 활용한 자동 진단 연구가 활발히 진행되고 있다.

1.2 관련 연구 및 문제 제기

최근 Cho et al.은 국내 10개 대학병원에서 수집된 43,338건의 대규모 데이터셋을 바탕으로 ResNet50 기반의 다중 분류 모델을 개발하였다 [2][3]. 해당 연구는 영상 데이터와 함께 재태 주수 (Gestational Age, GA) 및 출생 체중(Birth Weight, BW) 정보를 입력단에서 채널로 결합하는 Early Fusion 방식을 제안하여 진단 정확도를 높였다 [2].

그러나 이 방식은 단일 수치인 임상 정보를 고해상도 이미지 크기(512 × 512)로 확장하여 결합하므로 메모리 효율성이 떨어지며, 무엇보다 임상 정보가 영상의 국소적 특징(Local Feature) 추출에 어떻게 기여하는지 모델링하지 못한다는 한계가 있다. 이는 TTN과 같이 미세한 병변 차이를 보이는 질환의 분류 성능 한계(F1-score 70.84%)로 이어졌다 [2].

최근에는 신생아 CXR의 낮은 화질이 진단 및 모델 성능의 주요 병목으로 지적되면서, 이미지 품질을 향상시키는 전처리 연구도 활발히 진행되고 있다. 예를 들어, Han et al.은 Bayesian Optimization 기반 CLAHE(BO-CLAHE)를 적용하여 신생아 CXR 대비를 개선하고 TTN 분류 성능을 유의하게 향상시켰다고 보고하였다 [4].

1.3 본 연구의 기여

본 연구에서는 임상 정보가 영상 특징의 중요도를 결정하는 가이드스(Guidance) 역할을 수행해야 한다는 가설 하에, Clinical-Context Gated Network (CCGN)를 제안한다. 주요 기여는 다음과 같다.

1. 임상 정보(Context)를 키(Key)로 사용하여 영상 특징 맵을 Gating하는 새로운 융합 아키텍처를 제안한다.
2. 기존의 단순 결합 방식(Early/Late Fusion) 뿐만 아니라, 최신 어텐션 기법(SE-Net [5], CBAM [6])과의 비교 실험을 통해 제안 모델의 공학적 우수성을 입증한다.
3. 임상적 사전 지식(Prior Knowledge)을 공학적으로 모델링하여, 영상학적 특징이 모호한 TTN 및 초기 RDS의 감별 진단 성능을 획기적으로 개선함으로써 모델의 임상적 유효성을 입증한다.

2. 제안 방법

2.1 전체 구조 (Overall Architecture)

제안 모델은 영상 특징을 추출하는 Visual Encoder, 임상 정보를 처리하는 Clinical Context Encoder, 그리고 두 정보를 융합하는 Gating Module로 구성된다.

1. **Visual Encoder:** ResNet50 [3]을 백본(Backbone)으로 사용하여, 입력 이미지 (I)로부터 특징 맵 $F_{vis} \in R^{(C \times H \times W)}$ 를 추출한다 (C=2048).
2. **Clinical Context Encoder:** GA와 BW는 정규화된 벡터 $v_{clin} \in R^2$ 로 입력된다

2.2 Clinical-Context Gating Module (CCGM)

본 연구의 핵심인 CCGM은 임상 정보에 따라 영상 특징 채널의 가중치를 동적으로 조절한다. 이는 마치 오디오 이퀄라이저(Equalizer)가 장르에 따라 특정 주파수를 조절하듯, 환자의 임상 문맥(Context)을 조건부 인자(Conditioning Factor)로 활용하여 진단에 불필요한 특징은 억제(Suppress)하고 결정적인 특징은 증폭(Excite)시키는 필터링 메커니즘이다.

1. **Context Embedding:** 임상 벡터 v_{clin} 은 다층 퍼셉트론(MLP)을 거쳐 중간 임베딩 z 로 변환된다.

$$z = ReLU(w_1 \cdot v_{clin}) \tag{1}$$

2. **Attention Generation:** 임베딩 z 는 다시 선형 변환과 시그모이드 함수를 통해 채널별 중요도 벡

터 $A_{gate} \in R^C$ 를 생성한다.

$$A_{gate} = \sigma(W_2 \cdot z) \quad (2)$$

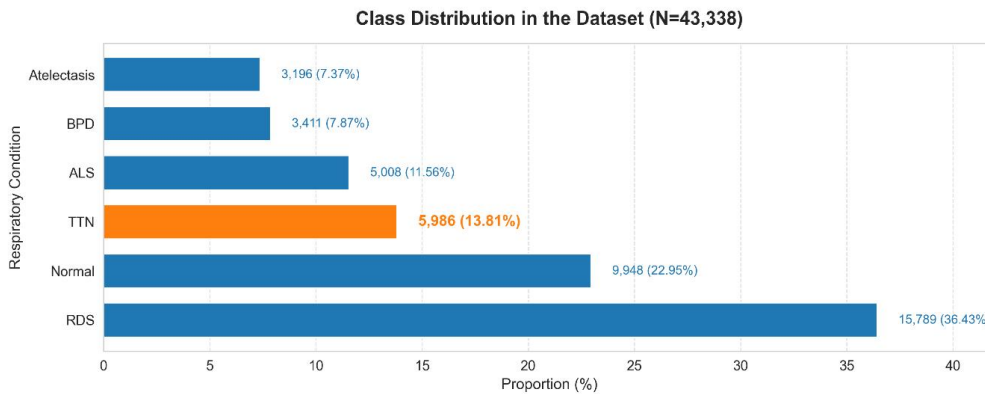
3. Feature Gating: 생성된 어텐션 벡터 A_{gate} 는 원본 영상 특징 맵 F_{vis} 와 채널별 곱(Channel-wise Multiplication) 연산을 수행한다.

$$F_{out} = F_{vis} \otimes A_{gate} \quad (3)$$

3. 실험 및 결과

3.1 데이터셋 및 전처리

본 연구는 선행 연구 [2] 데이터를 기반으로 11개 대학병원 데이터셋을 사용하였다. 총 43,338건의 CXR 이미지와 매칭된 임상 정보(GA, BW)가 포함되어 있다. 데이터셋의 클래스별 분포는 [그림 1]과 같다.



[그림 1] 데이터셋 클래스 분포

RDS가 가장 큰 비중(36.43%)을 차지. TTN(13.81%)은 상대적으로 적은 비중을 보여 클래스 불균형 나타남.

[Fig. 1] Class distribution of neonatal respiratory diseases in the multicenter chest radiograph dataset.

Respiratory distress syndrome (RDS) accounts for the largest proportion (36.43%), whereas transient tachypnea of the newborn (TTN) constitutes 13.81%, indicating a noticeable class imbalance.

모든 이미지는 Lung Mask Segmentation 후 512×512 픽셀로 크기를 조정(Resize)하였으며, 데이

터는 학습(34,598), 검증(4,370), 테스트(4,370) 셋으로 분할되었다. 훈련 단계에서는 좌우 반전 (horizontal flip), 밝기, 대비(brightness, contrast) 조절 등 기본적인 데이터 증강(data augmentation)을 적용하여 과적합(overfitting)을 완화하였다. 검증 및 테스트 데이터에는 증강을 적용하지 않고, 정규화(normalization)만 수행하였다.

모든 모델은 Adam optimizer를 사용하여 학습하였으며, 초기 학습률(learning rate)은 1×10^{-4} , weight decay는 1×10^{-5} 로 설정하였다. 배치 크기(batch size)는 32, 최대 100 epoch까지 학습하였고, 검증 손실(validation loss)이 개선되지 않을 경우 조기 종료(early stopping)를 적용하였다.

3.2 비교 실험 결과

제안 모델의 우수성을 검증하기 위해 ResNet50(Baseline), Early/Late Fusion, SE-ResNet50 [5], CBAM-ResNet50 [6], ConvNeXt-Tiny [7], Swin-T [8], 그리고 제안하는 CCGN까지 총 8가지 방법론을 [표 1]에서 비교 분석하였다. 클래스 불균형 및 다중 클래스 분류 성능을 공정하게 평가하기 위해, 데이터 불균형에 상대적으로 강건한 Macro-AUROC를 주요 지표로 사용하고, F1-score를 보조 지표로 제시하였다.

[표 1] 융합 방식 및 어텐션 기법에 따른 모델 간 성능 비교

[Table 1] Performance comparison according to fusion strategies and attention mechanisms

Methods	Fusion Type	F1-score	Macro-AUROC	Params (M)
Baseline (ResNet50)	Image Only	0.8368	0.9122	23.51
ConvNeXt-Tiny	Image Only	0.8334	0.9097	28.60
Swin-T	Image Only (Transformer)	0.8288	0.9054	29.00
Previous Work [2]	Early Fusion (InputConcat)	0.8632	0.9352	23.51
Late Fusion	Feature Concat	0.8456	0.9260	23.53
SE-ResNet50	Channel Attention	0.8475	0.9281	26.04
CBAM-ResNet50	Spatial + Channel	0.8502	0.9301	26.06
Proposed (CCGN)	Clinical-Context Gating	0.8728	0.9457	23.58

[표 1]에서 확인할 수 있듯이, Baseline ResNet50의 F1-score와 Macro-AUROC는 각각 0.8368, 0.9122로 나타났다. 선행 Early Fusion 기반 모델(Previous Work [2])은 영상과 임상 정보를 입력단에서 단순 결합함으로써 F1-score 0.8632, Macro-AUROC 0.9352까지 성능을 향상시켰다. 제안하는 CCGN은 동일한 데이터셋 및 실험 조건에서 F1-score 0.8728, Macro-AUROC 0.9457을 기록하여, Baseline 대비 약 3.6%p, 선행 Early Fusion 모델 대비 약 1.0%p의 F1-score 향상과 함께 AUROC 측면에서도 일관된 개선을 보였다.

ConvNeXt-Tiny와 Swin-T와 같은 최근 제안된 일반 목적 백본(backbone)들은 이미지 기반 분류에

서는 경쟁력 있는 구조이지만, 본 연구의 신생아 CXR 데이터셋에서는 F1-score와 Macro-AUROC 모두에서 Baseline ResNet50과 유사하거나 소폭 낮은 성능을 보였다. 이는 제한된 데이터 규모와 도메인 특성 상, 복잡한 일반 목적 백본보다 적절한 귀납적 편향(inductive bias)을 지닌 ResNet 계열 구조에 임상 문맥(context)을 주입하는 방식이 더 효율적일 수 있음을 시사한다.

또한, CCGN은 ResNet50 백본 대비 약 0.07M(23.51M → 23.58M, 약 0.3%) 수준의 매우 제한적인 파라미터 증가만을 수반하면서도 가장 높은 F1-score와 Macro-AUROC를 기록하였다. 이러한 결과는 단순히 채널/공간 어텐션 모듈을 추가하는 것보다, 임상 정보를 조건부 게이팅 신호로 활용하는 것이 신생아 호흡기 질환 분류 문제에 더 효과적임을 보여준다.

3.3 질환별 성능 분석

[표 2]는 전체 6개 클래스에 대한 F1-score를 비교한 결과로, 제안 모델(CCGN)이 대부분의 클래스에서 향상을 보였음을 확인할 수 있다. 특히 기존 연구에서 가장 낮은 성능을 보였던 TTN은 70.84%에서 77.60%로 크게 향상되었으며, Normal, RDS, Atelectasis 등에서도 안정적인 개선이 관찰되었다.

[표 2] 선행 Early Fusion 모델과 제안한 CCGN 간 질환별 F1-score 비교

[Table 2] Per-class F1-score comparison between the previous early-fusion model and the proposed CCGN

Class	Previous Work (Early Fusion)	Proposed (CCGN)	Improvement
BPD	0.9219	0.9269	0.51%
ALS	0.9065	0.9045	-0.20%
Atelectasis	0.8656	0.8721	0.64%
Normal	0.8738	0.8852	1.21%
RDS	0.9030	0.9177	1.49%
TTN	0.7084	0.7760	6.76%

4. 고찰

4.1. 임상 문맥 기반 게이팅(Gating)의 효과 분석

본 연구의 핵심 가설은 ‘임상 정보(재태 주수, 출생 체중)가 영상 특징 추출의 가이드라인(Guidance) 역할을 수행해야 한다’는 것이다. 실험 결과, 제안하는 CCGN(Clinical-Context Gated Network) 모델은 단순 결합 방식(Early Fusion) 대비 F1-score와 AUROC 모두에서 우수한 성능을 보였다.

이는 서론에서 언급한 정보 밀도의 불균형(Information Density Imbalance)에 따른 정보 희석

(dilution of information) 문제를 Gating Module이 미숙아(Preterm)와 만삭아(Term)의 영상적 특징을 구별하는 ‘필터(Filter)’로 작동함으로써 완화된 결과로 해석된다.

특히 주목할 점은, CCGN이 선형 Early Fusion 모델에 비해 F1-score 관점에서는 약 1.0%p 수준의 비교적 제한적인 개선을 보였음에도 불구하고, Macro-AUROC에서는 보다 뚜렷한 성능 향상을 달성했다는 점이다. 이는 단순히 임계값(threshold) 기준의 분류 정확도만 개선된 것이 아니라, 전체 score 분포 전반에서 양성/음성 클래스의 분리가 향상되었음을 의미한다. 다시 말해, CCGN이 임상 정보를 활용해 feature space 상에서 질환 간 경계를 보다 명확하게 재구성하고 있음을 시사한다. 임상 문맥 기반 게이팅이 ‘정답에 가까운 샘플만 더 잘 맞추는 것’이 아니라, 전 구간에서 decision boundary를 안정적으로 이동시키는 역할을 수행한다는 점에서, 실제 임상 적용 시 false positive와 false negative를 동시에 줄이는 데 기여할 수 있을 것으로 기대된다.

4.2. TTN 분류 성능 향상의 임상적 의의

가장 고무적인 성과는 신생아 일과성 빈호흡(TTN)의 분류 성능, 특히 F1-score의 뚜렷한 향상이다. 선형 연구에서 TTN은 RDS나 정상 폐 소견과 혼동되어 질환들 중 가장 낮은 F1-score 70.84%를 보였다. 이는 TTN이 영상학적으로 경미한 폐부종이나 폐문부의 선상 음영 등 미세한 특징만을 보이기 때문이다.

본 연구의 CCGN 모델은 TTN의 F1-score를 70.84%에서 77.60%로 약 6.8%p 향상시켰다. 이는 재태 주수 정보가 모델에 강력한 사전 확률(Prior Probability)로 작용하여, 만삭아에서 주로 발생하는 TTN의 특징을 더욱 예민하게 포착하게 만든 결과이다. 이러한 결과는 실제 NICU 임상 현장에서 감별 진단의 보조 도구(CDSS)로서의 효용성을 크게 높일 수 있음을 시사한다.

4.3. 기존 어텐션 기법 및 최신 백본 아키텍처와의 비교

본 연구에서는 영상 내부 정보만을 활용하는 SE-Net과 CBAM 기반 모델과도 성능을 비교하였다. 이들 어텐션 모듈은 Baseline 대비 일정 수준의 성능 향상은 제공하지만, Early Fusion이나 제안한 CCGN에서 관찰된 정도의 이득에는 미치지 못했다. 이는 신생아 CXR와 같이 영상 패턴이 단순하지만 클래스 간 경계가 모호한 문제에서는, 이미지 내부의 중요 영역을 재가중(re-weighting)하는 것만으로는 충분하지 않으며, 외부 임상 정보를 함께 고려하는 것이 중요함을 시사한다.

ConvNeXt-Tiny와 Swin-T와 같은 최근 제안된 일반 목적 백본(backbone) 역시 실험에 포함하였다. 이들 모델은 대규모 자연 이미지 데이터셋을 대상으로 설계된 구조로, 이론적으로는 더 강력한 표현력을 지니지만, 본 연구의 NICU CXR 데이터셋에서는 Baseline ResNet50과 유사하거나 다소 낮은 F1-score 및 Macro-AUROC를 보였다. 이는 (1) 제한된 데이터 규모, (2) 상대적으로 단순한 흉부

X-ray 패턴, (3) 임상 정보 미활용과 같은 요인이 복합적으로 작용한 결과로 해석할 수 있다.

반면, CCGN은 ResNet50 수준의 구조적 단순성을 유지한 채, 임상 정보를 채널 게이팅 신호로 주입함으로써 영상-임상 간 상호작용을 명시적으로 모델링하였다. 그 결과, 일반 목적 백본이나 이미지 기반 어텐션 모듈만을 사용하는 경우보다 더 높은 F1-score와 Macro-AUROC를 달성하였다. 이러한 결과는 의료 영상 진단 문제에서 ‘모델을 더 복잡하게 만드는 것’보다, 도메인 특화 임상 문맥(clinical context)을 어떻게 설계에 반영하느냐가 성능 개선에 더 결정적일 수 있음을 보여준다.

5. 결론

본 연구에서는 신생아 호흡기 질환 진단의 정확도를 높이기 위해, 임상 정보를 활용하여 영상 특징을 동적으로 제어하는 Clinical-Context Gated Network (CCGN)를 제안하였다. 제안 모델은 11개 대학병원의 대규모 데이터셋을 통해 검증되었으며, 기존의 단순 결합 방식(Early/Late Fusion)보다 효율적이고 강력한 분류 성능을 입증하였다.

본 연구의 주요 기여는 다음과 같다. CCGN은 임상 정보를 단순 입력 변수가 아닌 영상 해석을 위한 ‘주의 집중(Attention) 키’로 활용하는 새로운 융합 구조를 제시하고, 다중 클래스 분류에서 Macro-AUROC를 통해 모델의 강건함을 평가함으로써 기존 연구의 난제였던 TTN과 초기 RDS 감별 성능을 유의하게 개선하였다.

결론적으로, 본 연구에서 제안한 CCGN 모델은 신생아 중환자실(NICU) 의료진의 의사결정을 효과적으로 보조할 수 있는 잠재력을 확인하였다. 향후에는 시계열 영상 분석 및 설명 가능한 AI(XAI) 기술을 접목하여, 진단의 정확성뿐만 아니라 임상적 해석 가능성을 높이는 방향으로 연구를 지속할 계획이다.

References

- [1] C. L. Hermansen, K. N. Lorah, "Respiratory distress in the newborn", *American Family Physician*, vol. 76, no. 7, October 2007, pp. 987-994.
- [2] H. W. Cho, S. Jung, K. H. Park, J. W. Choi, J. S. Heo, J. Kim, H. Yun, D. Yu, J. Son, B. M. Choi, "Deep-learning-based multi-class classification for neonatal respiratory diseases on chest radiographs in neonatal intensive care units", *Neonatology*, vol. 122, no. 4, March 2025, pp. 446-454, doi: 10.1159/000545107.
- [3] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [4] J. Han, B. Choi, J. Y. Kim, Y. Lee, "BO-CLAHE enhancing neonatal chest X-ray image quality for improved lesion classification", *Scientific Reports*, vol. 15, no. 4931, January 2025, doi: 10.1038/s41598-025-86911-1.
- [5] J. Hu, L. Shen, G. Sun, "Squeeze-and-Excitation Networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 18-22, 2018, Salt Lake City, UT, USA, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.
- [6] S. Woo, J. Park, J. Y. Lee, I. S. Kweon, "CBAM: Convolutional Block Attention Module", *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 11211, September 8-14, 2018, Munich, Germany, pp. 3-19, doi: 10.1007/978-3-030-01234-2_1.
- [7] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, "A ConvNet for the 2020s", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 19-24, 2022, New Orleans, LA, USA, pp. 11976-11986, doi: 10.1109/CVPR52688.2022.01168.
- [8] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 11-17, 2021, Montreal, QC, Canada, pp. 9992-10002, doi: 10.1109/ICCV48922.2021.00986.