일반논문 (Regular Paper)
방송공학회논문지 제30권 제3호, 2025년 5월 (JBE Vol.30, No.3, May 2025)
https://doi.org/10.5909/JBE.2025.30.3.449
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

태권도 시합 판정 보조를 위한 가변 초점 변조 기반 겨루기 동작 인식 시스템

이 하 랑[a)], 윤 경 로[b)], 심 영 균[c)‡]

Adaptive Focal Modulation-Based Kyorugi Action Recognition System for Taekwondo Match Judgment Assistance

Harang Lee[a)], Kyoungro Yoon[a)], and Youngkyun Sim[c)‡]

요 약

본 논문은 태권도 겨루기 경기의 공정성과 신뢰성 문제를 해결하기 위하여 인공지능 기반의 판정 보조 시스템을 제안하며, 특히 태권도 동작 인식을 위하여 가변 초점 변조 기반 동작 인식 모델을 제안하였다. 기존의 판정 시스템은 정상적인 태권도 동작으로 인정되지 못하는 변칙 기술을 통한 부정 득점을 효과적으로 방지하지 못하는 한계가 있었다. 이를 개선하기 위해 본 논문에서는 객체 탐지 알고리즘을 활용하여 경기 내 선수의 동작을 집중적으로 분석하고, 인공지능 모델을 이용하여 시공간적 정보를 독립적으로 처리하는 새로운 접근 방식을 도입하였다. 또한, 프레임의 선형 결합을 통한 노이즈를 부여하는 방식을 적용하여 배경 정보의 영향을 최소화하고, 주요 객체의 동작 변화에 집중하도록 만들었다. 제안된 모델은 기존 동작 인식 모델보다 높은 성능을 보였으며, 다양한 태권도 데이터셋에서 변칙 발차기 탐지 성능이 개선됨을 확인하였다.

Abstract

This paper proposes an adaptive focal modulation-based motion recognition model to enhance the fairness and reliability of judgement made in Taekwondo sparring competitions. Existing judgment systems have limitations in preventing unfair scoring through irregular techniques. To address these issues, this study introduces an object detection algorithm that focuses on analyzing competitor movements, coupled with an AI-driven approach for independently processing spatio-temporal information. Additionally, a frame-wise linear combination strategy is applied to introduce noise, minimizing the influence of background elements and improving the model's focus on motion variations. The proposed model demonstrates superior performance compared to conventional action recognition methods and exhibits enhanced detection accuracy for irregular kicking techniques across diverse Taekwondo datasets.

Keyword : Taekwondo, Judgment Assistance, Action Recognition, Object Detection, Focal Modulation

Ⅰ. 서 론

4차 산업혁명과 함께 다양한 정보 통신 기술이 발전하며 스포츠 분야에서 전문적인 분석과 공정한 판정을 위해 IT 기술들을 적극적으로 도입하고 있다. 이러한 흐름에 따라 태권도 분야에서도 오심으로 인한 공정성 논란 해소를 위해 다양한 기술을 적용하고 있다. 태권도 겨루기는 단순히 심판의 관찰로 타격 부위와 타격 강도를 판단하는 방식으로 판정했기 때문에 심판의 주관성 개입 및 일관성 부족으로 인한 공정성 문제가 제기되었고, 이로 인해 올림픽 경기에서 퇴출 가능성이 거론되기도 하였다.

이러한 공정성 문제를 해결하고자 2009년에는 헤드 기어 및 몸통 보호대 내에 충격 감지 시스템을 갖추어 타격의 유효를 결정하고 이를 무선 데이터 통신 시스템을 통해 점수를 표출하는 전자 호구 기반 판정 시스템을 도입하여 판정의 신뢰도를 높였다[1]. 이와 함께 판정에 대한 공정성 향상을 위해 2009년 세계 태권도 선수권 대회에서 영상 판독 시스템을 사용하였다[2]. 이때, 객관성을 높이기 위하여 판정용 카메라를 전후좌우뿐만 아니라 천장에도 설치하고 이를 관중도 함께 확인할 수 있도록 경기장 내 대형 스크린을 통해 송출하는 방식을 채택하였다. 또한 영상 판독 시스템의 긴 지연 시간 해결을 위해 2020년 도쿄 올림픽에서 4D 카메라를 도입하여 영상 판독 시간을 5초 이내로 단축시켰다[3].

이러한 공정성을 확보하기 위한 노력에도 불구하고 여전히 오심으로 인한 문제가 존재한다. 태권도 경기 규칙 제21조 비디오 판독을 살펴보면, 코치가 비디오 판독권이 없는 상황에서라도 3회전 종료 전 10초 이내 또는 골든포인트 회전 도중 부심 중 누구라도 득점에 이의가 있을 경우 비디오 판독을 요청할 수 있다[4]. 하지만 2018년 한국 중고등학교 태권도 연맹 회장기 전국 태권도 대회에서 경기 규칙을 숙지하지 못한 영상 판독관이 영상 판독을 거부했다[5].

전자 호구 시스템은 태권도 경기의 흐름을 크게 변화시켰다. 전자 호구의 등장으로 인해 강한 힘으로 상대 선수를 타격하는 선제 공격 위주의 경기에서 소극적이고 방어적인 동작 위주의 경기로 변화하였다[6]. 태권도 협회는 경기 규칙 개정을 통해 몸통 앞발 공격을 줄이고자 하는 등 많은 변화를 주며 노력해 왔지만 전자 호구는 유효 타격을 단순히 타격 부위나 강도로 결정하기 때문에 규칙 개정만으로 해결할 수 없는 문제가 남아있다. 먼저, 센서가 부착되지 않은 부분을 타격하는 경우 득점 인정이 되지 않거나 호구 자체에 센서 오류나 신호 전달 오류가 있는 경우도 존재한다[7][8]. 또한 변칙 발차기로 센서가 부착된 부위를 타격하는 경우에도 유효 득점이 되어 이를 이용해 부정 득점을 얻는 문제점도 존재한다[9]. 전자 호구 시스템 도입은 심판의 주관성 개입을 어느 정도 해소하였지만 새로운 공정성 문제를 만들어내고 있다[10].

본 논문에서는 태권도 겨루기 경기의 공정성 및 신뢰성 문제를 해결하기 위하여 실시간 경기에서 변칙 동작과 정상 동작을 구분할 수 있는 인공지능 행동 인식 시스템을 제안한다. 태권도 겨루기는 태권도 품새나 격파와 달리 다양한 변수가 존재한다. 특히 전자 호구 시스템 도입 이후 다양한 변칙 기술이 등장했지만, 기존 시스템은 변칙 동작이 득점으로 이어지는 것을 제대로 막지 못하고 있다. 실시간 경기에서는 빠르고 정확하게 동작을 판별하여 유효 득점을 결정해야 하지만 많은 동작이 빠르게 이루어지는 겨루기에서 사람이 모든 동작을 정확하게 판별하기엔 역부족이다. 따라서 본 연구에서는 인공지능 기반 동작 인식을 위해 태권도 데이터셋을 구축하고, 빠른 인식을 위한 가변 초점 변조 방식을 제안한다. 이러한 동작 인식 모델을 실제 경기에 적용한다면 영상 판독 시스템과 전자 호구 시스템 도입에 이어 태권도 겨루기 경기의 공정성뿐만 아니라 흥미를 높일 수 있는 도구가 될 것으로 기대한다.

a) 건국대학교 스마트ICT융합공학과(Dept. of Smart ICT Convergence, Konkuk University)
b) 건국대학교 컴퓨터공학부(Dept. of Computer Science and Engineering, Koukuk University)
c) 우석대학교 체육복지융합연구소(Woosuk University)
‡ Corresponding Author : 심영균(Youngkyun Sim)
E-mail: simco76@naver.com
ORCID: https://orcid.org/0000-0001-8249-9121
※ 이 논문은 2023년 세계태권도연맹과 문화체육관광부의 지원을 받아 수행된 연구입니다. (WT-AI2023001)
· Manuscript April 15, 2025; Revised April 21, 2025; Accepted April 21, 2025.

Ⅱ. 관련 연구

1. 태권도 발차기 분류 연구

변칙 발차기와 정상 발차기를 구분하지 못하는 전자 호구의 단점을 이용한 부정 득점을 해결하기 위해 다양한 인공지능 모델을 이용한 변칙 발차기 분류 연구가 이루어지고 있다. 2020년에는 시계열 모델을 이용한 변칙 발차기 분류 모델이 제안되었다[11]. 이 연구는 전자 호구에 IMU (Inertia Measurement Unit) 센서를 부착하여 근전도, 각속도, 가속도, 지자계의 특징값을 기록하였다. 이렇게 주어진 데이터를 분류하기 위해 연속된 데이터에 대한 장기 의존성 문제를 완화한 LSTM 모델을 사용하였다[12]. 하지만 영상이나 이미지 없이 단순히 수치를 이용하였을 뿐만 아니라 수집 데이터가 적어 11개의 발차기에 대해 27.27%라는 매우 낮은 분류 정확도를 보여주었다.

2021년에는 기계 학습을 이용한 변칙 발차기 탐지 연구가 진행되었다[13]. 이 연구에서 태권도 선수들에게 4개의 센서를 부착하여 가속도 X, Y, Z축 값과 합성 가속도, 각속도 X, Y, Z축 값과 합성 각속도를 데이터로 수집하였다. 모델은 유사도 추정을 이용한 K-최근접 이웃(K-Nearest Neighbor)과 최적의 결정 경계를 구하는 서포트 벡터 머신(Support Vector Machine)을 사용하였다[14][15].

분류 정확도는 K-최근접 이웃이 89.83%, 서포트 벡터 머신이 91.39%로 높은 정확도를 보여주었지만, 변칙 발차기를 제기차기 형태로 발을 뻗어 차는 몽키킥 하나로 정의하였다. 따라서 단순히 일반 발차기와 변칙 발차기 2개의 클래스에 대해 분류 모델을 적용하였기 때문에 실제 경기에 적용하기에는 어렵다는 한계점이 존재한다.

2. 객체 탐지 연구

R-CNN은 선택적 검색(Selective Search) 기법을 활용한 객체 탐지 알고리즘이다[16]. 선택적 검색(Selective Search) 기법을 활용해 객체가 존재할 가능성이 높은 후보 영역(Region Proposal)을 생성하는 방식을 사용한다. 이미지를 여러 개의 작은 영역(Superpixel)으로 분할한 뒤, 계층적 통합 방식(Hierarchical Grouping Algorithm)을 적용해 유사한 영역들을 통합하는 과정으로 진행된다. 도출한 최대 2,000개의 후보 영역은 Alexnet을 이용하여 특징을 추출하고, 서포트 벡터 머신을 통해 분류를 진행한다. 마지막 단계에서는 경계 상자 회귀(Regression)를 통해 예측된 경계 상자와 실제 경계 상자 간의 차이를 최소화하도록 보정하는 과정을 수행한다. 이러한 방식은 높은 정확도를 갖지만, 후보 영역을 개별적으로 합성곱(Convolution) 연산을 수행해야 하므로 처리 속도가 느리고 메모리 사용량이 많다는 단점이 있다.

Yolo는 이미지 전체를 한 번에 처리하여 비교적 빠른 속도로 객체를 탐지할 수 있다[17]. 입력 이미지에 대해 사전에 정의된 그리드로 나누고 그리드 셀 내에 객체의 중심이 존재하는 경우에 경계 상자를 검출하고 신뢰도 점수를 예측한다. 이때, 비최대 억제(Non-maximum Suppression) 알고리즘을 사용해 중복된 경계 상자를 제거하는 과정을 거친다. 신뢰도 점수가 가장 높은 경계 상자를 기준으로 IoU(Intersection over Union)가 기준값 이상인 경계 상자를 제거하여 클래스별로 정확한 경계 상자를 남긴다. 이 방식은 클래스 분류와 경계 상자 예측을 동시에 수행하기 때문에 빠른 속도로 연산을 처리하여 실시간 환경에서 사용하기에 적합하다.

3. 동작 인식 연구

합성곱 신경망(Convolutional Neural Networks)은 초기에는 주로 이미지 데이터를 처리하는 데 사용되었지만, 이후 비디오 데이터로 연구가 확장되었다[18]. 초기 연구에서는 이미지에 사용되는 2차원 합성곱을 변형해 비디오 분류 문제에 적용하는 방식을 사용했다[19][20]. 비디오 데이터는 높은 차원을 가지고 다양한 시공간 정보를 포함하며 프레임 간 복잡한 의존성을 가진다. 하지만 2차원 합성곱은 연산 비용이 크지 않지만, 비디오 데이터의 시간 정보를 반영하는 데 한계가 존재한다. 이러한 한계점을 극복하기 위해서 H×W×C 형태에서 H×W×T×C 형태로 차원을 확장한 3차원 합성곱을 이용한 다양한 비디오 분류 연구들이 등장했다.

Slowfast는 비디오의 빠른 변화와 느린 변화를 동시에 학습하기 위해 이중 스트림(Two-Stream) 구조로 설계된 모델이다[21]. 이 모델은 슬로우 패스(Slow Path)와 패스트 패스(Fast Path)로 구성되어 공간적 정보와 시간적 정보를 각각 학습한다. 슬로우 패스는 낮은 주파수인 공간적인 정보에

더 집중하고 패스트 패스는 높은 주파수인 시간적인 정보에 더 집중하게 된다. 또한 3차원 합성곱을 적용하여 비디오의 시공간적 특징을 효과적으로 학습할 수 있어 세밀한 움직임을 정확하게 포착할 수 있다.

ViViT는 자연어 처리 분야에서 좋은 성능을 보여준 트랜스포머(Transformer)를 비디오 분야로 확장한 모델이다[22]. 비디오 데이터를 여러 개의 패치로 분할한 뒤, 위치 정보 보존을 위해 위치 임베딩(Position Embedding)을 사용한다. 이후 시공간 튜브(Spatio-Temporal Tube)를 활용해 멀티 헤드 셀프 어텐션(Multi-Head Self-Attention) 연산을 수행하여 중요도를 파악하고 시공간 정보를 동시에 처리한다는 장점이 있다. 또한 기존에 ViT(Vision Transformer)는 약한 유도 편향(Inductive bias)으로 많은 데이터가 필요하다는 단점이 있었다. 따라서 ViViT는 이러한 문제를 해결하기 위해 사전 학습된 ViT를 인플레이션하여 학습에 사용하였다.

MViT는 비디오 데이터의 3D 패치를 활용해 시공간 정보를 효과적으로 학습한다[23]. ViViT와는 달리 다중 스케일 피라미드(Multiscale Pyramid) 구조를 사용하여 공간 해상도를 줄이며 채널 용량을 확장하는 방식으로 다양한 해상도의 특징을 포착할 수 있도록 설계되었다. 고해상도에서는 세밀한 특징을, 저해상도에서는 큰 문맥 정보를 추출하여 다양한 정보를 고려할 수 있다.

VideoFocalNet은 초점 변조(Focal Modulation) 방식을 도입하여 지역적 문맥과 전역적 문맥을 모두 고려하는 방식을 사용했다[24]. 기존 셀프 어텐션(Self-Attention) 방식이 상호작용(Interaction) 단계 이후 집계(Aggregation) 단계를 수행했던 것과는 달리 VideoFocalNet은 시간적 정보와 공간적 정보를 집계한 뒤 적응적으로 변조하여 상호작용한다. 또한 시간적 정보를 포착하기 위해 1차원 합성곱을 사용하여 적은 계산 비용으로 복잡한 시간적 관계를 포착한다.

Ⅲ. 데이터 및 학습 모델

1. 데이터 수집 및 분석

영상 기반 동작 분류 모델 학습을 위해 세계 태권도 연맹에서 태권도 겨루기 경기 영상을 제공받아 직접 영상 클리핑(Clipping)을 수행하였다. 하지만 태권도 겨루기 경기 영상에서 수집한 데이터는 정상 발차기에 비해 변칙 발차기의 개수가 매우 부족하여 학습에 사용하기에 적합하지 않아 추가적인 촬영을 진행했다. 변칙 발차기의 비중을 높여 고등학교 체육관에서 겨루기 연습 경기를 진행하였으며, 이를 직접 촬영하여 데이터를 추가로 수집하고 클리핑하였다. 학습에는 각 동작 별로 60fps의 32개의 프레임으로 구성된 총 1,367개의 클립(Clip)을 사용했다. 자체적으로 일차 분류를 수행한 뒤, 각각의 클립에 대해 두 차례 검수를 수행하고, 마지막으로 태권도 전문가 검수까지 총 3번의 검수를 거쳐 데이터에 대한 정확도 및 신뢰성을 높이고자 하였다. 동작은 머리 부위를 차는 발차기 3가지와 몸통 부위를 차는 발차기 4가지로 분류하였다. 머리 발차기 데이터셋은 정상 발차기, 안 차기, 낚아 차기로 정의하였으며, 개수는 각각 164개, 380개, 273개이다. 이때 정상 발차기는 변칙 발차기 이외의 상대 머리 부위를 차는 동작으로 정의하였고 안 차기는 손으로 상대 선수를 잡거나 클린치 상황일 때 발 안쪽으로 상대 머리 부위를 차는 동작으로 정의하였다. 마지막으로 낚아 차기는 안 차기와 마찬가지로 손으로 상대를 잡거나 클린치 상황에서 이루어지는 발차기이다. 하지만 낚아 차기는 몸통과 발차기를 하는 다리의 각도가 90도 이상이 되어 다른 변칙 동작이나 정상 발차기와 명확히 구분된다. 몸통 발차기 데이터셋은 정상 발차기, 안 차기, 안 돌려차기 및 밖 돌려차기, 밀어차기로 정의하였으며, 개수는 각각 243개, 182개, 322개, 277개이다. 이때 변칙 발차기 이외의 몸통 부위를 차는 동작을 정상 발차기로 정의하였고 안 차기는 손으로 상대 선수를 잡거나 클린치 상황에서 발 안쪽으로 상대 몸통 부위를 차는 동작으로 정의하였다. 안 돌려차기 및 밖 돌려차기는 안 차기와 같은 상황에서 이루지는 발차기이지만 무릎을 안쪽이나 바깥쪽으로 구부려 몸통 부위를 찬다는 차이점이 있다. 마지막으로 밀어 차기는 다리를 쭉 뻗어 상대 선수의 몸통 부위를 발바닥으로 차는 동작을 말한다.

세계 태권도 경기 데이터는 대부분 배경이 변화하는 영상이며 배경에는 경기 정보를 표출하는 전광판이 포함되어 있다. 또한 경기에 참여하는 선수뿐만 아니라 심판과 관중을 포함하며 일부 영상은 중계 영상과 같이 점수표와 타격

부위 및 타격 강도가 표시된다. 직접 촬영한 데이터는 카메라를 고정한 상태로 촬영된 배경의 변화가 없는 영상이며, 심판을 포함하지 않는다. 일부 영상은 뒤쪽에 소수의 코치나 관중을 포함하고 있지만, 경기에 참여하는 선수에 비해 객체의 크기가 매우 작다. 학습에 사용하는 데이터는 배경의 움직임이 없고 고정된 상태로 대상 객체만을 포함하는 것이 가장 이상적이다. 따라서 세계 태권도 경기에서 클리핑한 클립들은 학습 전 한번 더 검수를 거쳐 비교적 배경의 변화가 적은 영상만을 학습 데이터로 사용하였다. 또한 실제 태권도 겨루기 경기에서 다수의 관중이나 심판을 포함하는 것을 고려하여 선수 이외의 객체가 담긴 영상도 학습에 사용하였다. 또한 원 경기 영상은 1개 이상의 경기를 포함하는 여러 가지 동작이 뒤섞인 매우 긴 영상으로 학습에 적합하지 않았다. 따라서 이를 학습에 적합한 1초 미만의 영상으로 클리핑하기 위해 타격 시점을 기록할 수 있는 클리핑 프로그램을 직접 작성하였다. 이 프로그램을 이용하여 영상의 타격하는 시점을 기록하면 타격 시점의 앞부분과 뒷부분을 포함하여 총 32프레임으로 구성된 클립이 만들어진다. 또한 타격 시점 기록과 동작 기록을 함께 수행하도록 만들어 영상 클리핑과 분류를 하나의 단계로 통합하여 효율적인 데이터 처리를 가능하게 했다.

2. 객체 탐지

동작 인식의 정확도를 높이기 위해 경계 상자 집계 단계를 추가하여 관중이나 심판을 최대한 제외하였다. 이 단계는 다양한 정보를 포함한 비디오에서 목표 객체 이외의 배경 정보를 덜어내는 역할을 수행하여 동작 인식의 목표가 되는 객체에 집중할 수 있도록 만들어준다. 먼저, 경계 상자 집계를 위해 객체 탐지 알고리즘을 적용하여 사람 객체에 대한 탐지를 수행한다. 객체 탐지는 태권도 겨루기 경기에서 실시간으로 빠르게 동작하는 모델을 구현하기 위해 다른 모델에 비해 빠른 속도로 탐지가 가능한 Yolo 모델을 사용한다. Yolo는 다양한 버전으로 공개가 되었고 Yolov5 모델이 가장 활발히 사용된다. 하지만 성능 향상을 위한 연구가 지속되어 최근에는 Yolov5 보다 성능을 약 20% 개선한 Yolov8 모델이 발표되었다. 따라서 본 연구에서는 오류를 낮추고 높은 정확도를 달성하기 위해 Yolov8 모델을 채

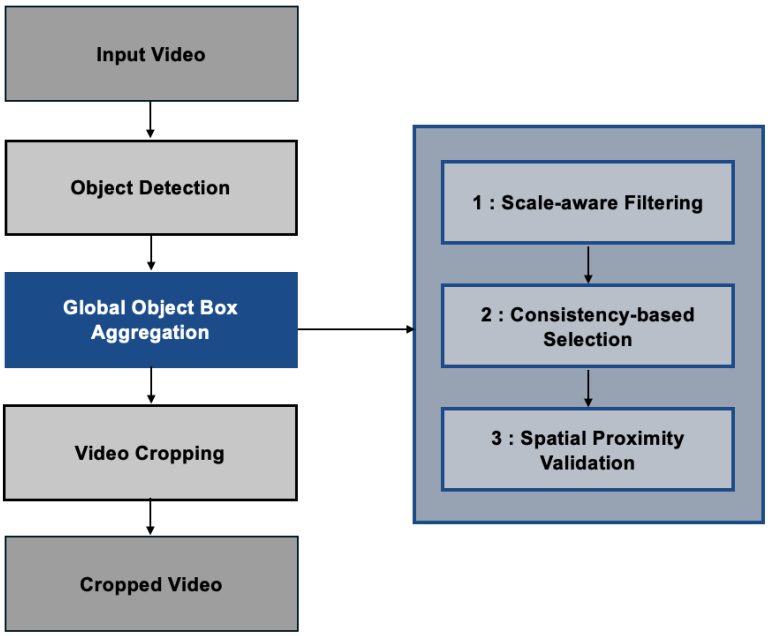

그림 1. 전역 객체 상자 집계 다이어그램
Fig. 1. Global object box aggregation process diagram

택하였다.

경계 상자 집계 단계에서는 비디오의 각 프레임을 객체 탐지 모델에 전달하여 사람 객체에 대한 경계 상자를 얻는다. 이후, 경계 상자를 집계하여 경계 상자 좌표의 최솟값과 최댓값을 이용해 비디오 크롭핑(Cropping)을 수행한다. 이를 위해 경계 상자 선택 알고리즘을 적용하여 원하는 객체만을 남길 수 있도록 설계하였다.

객체 탐지 단계에서 사용하는 Yolo는 중첩된 객체에 대하여 성능이 낮은 문제가 존재하지만, 본 연구에서는 선수를 명확히 구분하는 것보다는 선수 모두를 포함하는 장면을 크롭하는 것을 목적으로 이 모델은 사용한다. 따라서 중첩된 객체에 대해 경계 상자가 하나만 있더라도 선수 모두를 포함한 장면을 크롭할 수 있다. 또한 하나의 프레임이 아닌 여러 프레임에서 탐지된 경계 상자 정보를 집계하여 크롭하기 때문에 경계 상자를 탐지하지 못하는 경우를 보완할 수 있다.

경계 상자 선택 알고리즘은 크게 크기 기반 필터링과 위치 기반 필터링으로 나뉜다. 먼저, 크기 기반 필터링에서는 높이를 기준으로 경계 상자를 정렬하고, 기준치에 미달 되거나 초과하는 경계 상자는 제거한다. 필터링된 경계 상자는 순서대로 크기를 비교하여 그 차이가 기준치 이상인 경우 큰 크기의 경계 상자를 제거한다. 이 단계는 경기에 참여하는 선수보다 감독이 카메라 가까이에 위치하여 너무 크게 잡히는 경우, 심판에 대한 경계 상자를 제외할 수 있다.

위치 기반 필터링에서는 경계 상자 간의 근접도를 계산한다. 이를 통해 가장 근접해 있는 경계 상자들만 남기고 나머지 경계 상자는 제외한다. 이 방식은 멀리 떨어져 있는 관객을 제거하고 가장 가까운 위치에 있는 선수에 대한 경계 상자만을 남길 수 있다. 또한 크기 기반 필터링 단계에서 제거되지 않은 심판에 대한 경계 상자는 선수와 멀리 떨어져 있는 경우가 대부분이기 때문에 위치 기반 필터링 단계를 거쳐 제거할 수 있다.

선택된 경계 상자들은 마지막 전역 경계 상자 집계 단계를 거쳐 크롭핑을 위한 전역 경계 상자 좌표 정보를 얻는다. 선택된 경계 상자들에 대한 좌표를 집계하여 x축에 대한 최소 좌표 및 최대 좌표, y축에 대한 최소 좌표 및 최대 좌표를 얻어 총 4개의 좌표로 전역 경계 상자를 정의한다. 전역 경계 상자는 비디오 내에서 한 동작을 구성하는 모든 프레임에 걸쳐 선수들이 포함된 영역을 의미한다.

3. 동작 인식

동작 인식 단계에서는 경계 상자 집계 단계를 거쳐 크롭한 비디오 클립을 초점 변조 기반 동작 인식 모델의 입력으로 사용한다. 입력된 값은 집계 단계 이후 상호작용 단계를 거쳐 결과를 도출한다. 이러한 방식은 기존 셀프 어텐션 방식이 상호작용 단계 이후 집계 단계를 거쳤던 것과는 대조되게 반대되는 순서로 연산을 진행한다. 먼저 투영(Projection) 과정을 거쳐 얻은 게이트(Gate) 정보를 이용해 문맥별 반영 정도를 조절한다. 이렇게 조절된 정보는 집계 단계에서 통합되고, 채널 수를 조절하여 상호작용 단계로 전달된다. 상호작용 단계에서는 쿼리(Query)와 전달된 문

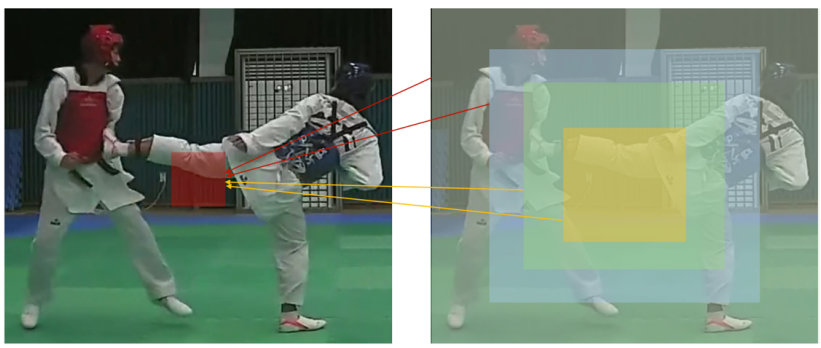

그림 2. 초점 변조 방식 쿼리 연산 시각화
Fig. 2. Visualization of focal modulation-based query computation

맥 정보를 요소별 곱(Element-wise Multiplication)을 수행하는데 이런 방식은 N이 토큰 개수이고 D가 차원 수일 때, 계산 복잡도가 O(N×D)가 된다. 기존 셀프 어텐션 방식은 쿼리와 키(Key)의 내적을 거치고, 소프트맥스(Softmax) 단계를 거쳐 마지막 단계에 밸류(Value)와 행렬 곱(Matrix Multiplication)을 수행하기 때문에 계산 복잡도가 $O(N^2 \times D)$가 된다. 따라서 초점 변조 기반 방식이 더 적은 연산량을 가지기 때문에 비디오와 같은 고차원 데이터를 더욱 효율적으로 처리할 수 있다.

기존 트랜스포머 기반 동작 인식 모델들은 비디오 데이터를 패치를 분할하고, 모든 패치에 대해 동일한 가중치를 적용하여 처리하는 방식을 적용하였다. 하지만 이러한 방식은 중요한 패치와 중요하지 않은 패치를 구분하지 않고 균등하게 패치를 처리하여 태권도 변칙 동작 구분과 같이 미세한 차이를 구분하는 작업에는 적합하지 않다. 하지만 초점 변조 기반 동작 인식 모델은 시간적 정보나 공간적 정보가 더 높은 중요도를 가지는 패치에 더 높은 가중치를 부여하여 차등적으로 처리하기 때문에 복잡하고 미세한 차이를 잘 구분한다. 이러한 방식은 불필요한 연산을 줄이고 문맥 정보를 효과적으로 활용하여 기존 트랜스포머 기반 모델들이 직면했던 높은 연산량 문제를 해결하고 높은 동작 인식 성능을 유지할 수 있도록 만들었다.

비디오 인식은 이미지 인식과 달리 시간 정보와 공간 정보를 모두 고려해야 할 필요성이 있다. 본 논문에서는 이러한 비디오 데이터의 특성을 고려하여 시간 정보와 공간 정보에 서로 다른 가중치를 주는 방식을 제안하여 모델이 동작 흐름에 대해 더 명확하게 인식할 수 있도록 만들었다. 기존 초점 변조 기반 동작 인식 모델은 문맥화, 집계, 변조 단계를 시간 정보와 공간 정보에 대해 각각 따로 처리한 뒤 상호작용하여 시공간 정보를 소실 없이 적절히 반영할 수 있다. 하지만 동작에 따라 서로 다른 공간 정보의 중요도와 시간 정보의 중요도를 반영하지 못한다는 단점이 있다.

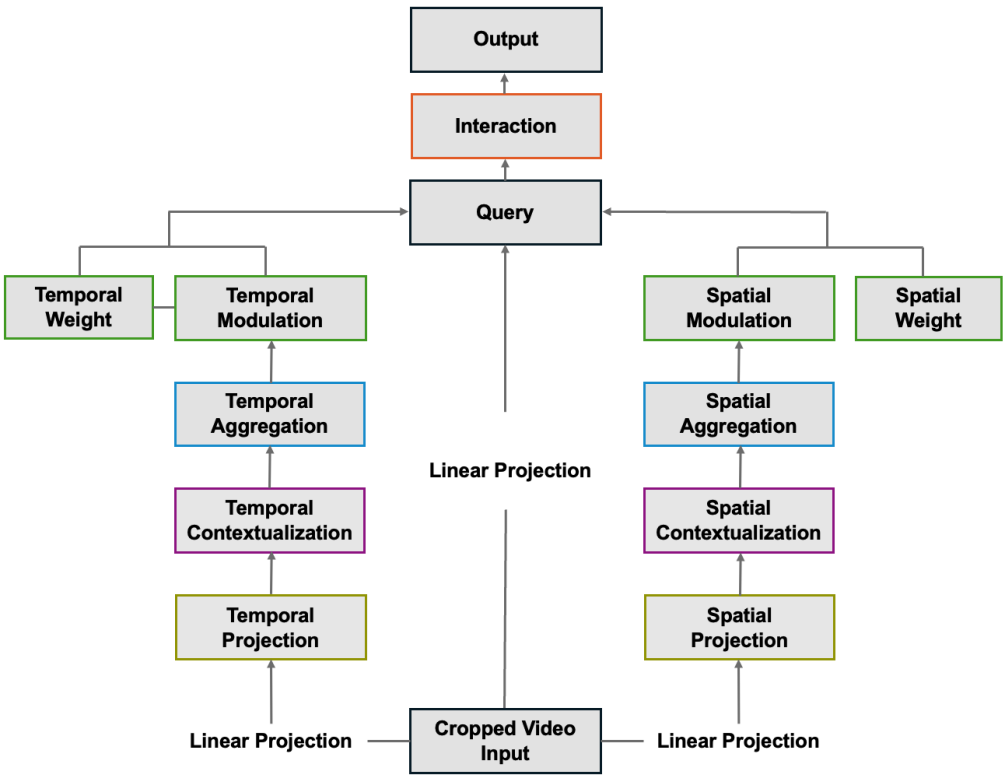

그림 3. VideoFocalNet 구조
Fig. 3. VideoFocalNet Architecture

태권도 정상 동작과 변칙 동작 간에는 시간적 차이가 큰 경우와 공간적 차이가 큰 경우가 나뉜다. 즉, 시간 정보와 공간 정보의 중요도를 서로 다르게 조절하여 쿼리에 적용해야 비디오 데이터의 동작 별 차이에 대해 더 효과적으로 반응할 수 있다. 하지만 시간 정보와 공간 정보의 중요도는 상충 관계를 갖지 않기 때문에 시간 모듈레이션과 공간 모듈레이션 각각에 따로 모듈레이션 가중치를 적용하는 것이 필요하다.

$$y_i = T_{st}(W_s \bullet M_s(i_t, X_{st,t}), W_t \bullet M_t(i_{hw}, X_{st,hw}), x_i) \quad (1)$$

특히 본 논문에서 제안하는 모듈레이션 가중치를 이용한 초점 변조 수식은 식 (1)과 같다. 수식에서 T는 시간적 정보와 공간적 정보의 상호작용을 의미하고, M_s은 특정 시간 t에 대한 공간적 문맥 정보를 나타내고 M_t은 특정 공간 s에 대한 시간적 문맥 정보를 나타낸다. 이때 공간 모듈레이션에서는 특정 시간에서의 공간 위치를 나타내는 인덱스 i_t와 특정 시간에서의 공간적 특징인 $X_{st,t}$를 사용한다. 시간 모듈레이션에서는 특정 공간에서 시간 위치를 나타내는 인덱스 i_s와 특정 공간에서의 시간적 특징인 $X_{st,hw}$를 사용한다. 마지막 단계에서는 공간 모듈레이션에는 가중치 W_s를 곱하고, 시간 모듈레이션에는 가중치 W_t를 곱하여 쿼리인 x_i와 상호작용한다.

$$y_i = q(x_i) \odot W_s \bullet m_s(i_t, X_{st,t}) \odot W_t \bullet m_t(i_{hw}, X_{st,hw}) \quad (2)$$

초점 변조 모듈레이션을 구체화한 수식은 식 (2)와 같다. 앞서 언급한 쿼리는 입력값인 x_i에 대해 선형 변환 q를 수행하여 얻을 수 있다. 또한 상호작용 단계에서는 쿼리와 모듈레이션 가중치를 곱한 시공간 정보를 요소별 곱셈을 통해 출력값을 얻는다.

Ⅳ. 결과 및 분석

동작 인식 시스템은 자체적으로 수집 및 생성한 태권도 영상 데이터에서 뒤쪽에서부터 원하는 개수의 프레임을 가져와 224×224 크기로 조정된 데이터를 이용하여 학습 및 평가를 진행한다. 데이터셋은 6:2:2 비율로 나누어 각각 학습, 검증, 테스트 데이터로 사용하였다. 불균형 데이터를 사용하기 때문에 학습 중 특정 클래스에 편중되어 학습되는 것을 방지하기 위해 데이터 분할 시 클래스 비율을 균등하게 맞추었다. 또한 모델의 일반화 성능 확보를 위해 이미지 반전 및 컬러 변환과 같은 데이터 증강 기법을 적용하여 학습을 진행하였다.

손실함수로는 교차 엔트로피 손실(Cross-Entropy Loss) 함수를 사용한다. 이는 다중 클래스 분류 문제에 사용되는 손실함수로 모델이 예측값이 정답에 가까울수록 손실을 낮게 계산한다. 교차 엔트로피 손실함수의 수식은 식 (3)과 같다.

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_i, c) \quad (3)$$

수식에서 N은 데이터의 개수, C는 클래스의 개수이며 y는 데이터에 대한 클래스를 나타내고 ŷ은 모델이 예측한 클래스에 대한 확률을 나타낸다.

성능 측정 지표로는 정밀도(Precision), 재현율(Recall), F1-점수(F1-score), 정확도(Accuracy)를 사용한다. 정밀도는 모델이 양성으로 예측한 샘플 중, 실제로 양성인 샘플의 비율이다. 정밀도는 식 (4)와 같이 정의된다.

$$\mathrm{Precision} = \frac{TP}{TP + FP} \quad (4)$$

재현율은 실제 양성 샘플 중에서 모델이 양성이라고 예측한 비율이다. 재현율은 식 (5)와 같이 정의된다.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F1-점수는 정밀도와 재현율의 조화 평균(harmonic mean)으로 두 지표를 동시에 고려할 수 있다. F1-점수는 식 (6)과 같이 정의된다.

$$F1 - score = 2 \times \frac{\mathrm{Precision} \times Recall}{\mathrm{Precision} + Recall} \quad (6)$$

정확도는 전체 데이터 중 모델이 답을 맞힌 샘플의 비율을 말한다. 이러한 지표는 직관적이며 이해하기 쉽다는 장점이 있다. 정확도는 식 (7)과 같이 정의된다.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

TP는 모델이 하나의 클래스에 대해서 긍정으로 정확히 예측한 경우를 말하며 FP는 실제로는 부정인 것을 모델이 긍정으로 잘못 예측한 경우를 말한다. TN은 모델이 하나의 클래스에 대해 부정으로 정확히 예측한 경우이고, FN은 실제로는 긍정인 것을 모델이 부정으로 잘못 예측한 경우를 말한다.

모듈레이션값과 쿼리를 곱하기 전에 공간적 모듈레이션 정보와 시간적 모듈레이션 정보를 적절히 조절할 수 있도록 모듈레이션 파라미터를 적용 시 기존 모델과의 성능을 비교하였다. 이 경우 머리 발차기 데이터셋의 경우에는 최대 풀링을 적용하였을 때 기존 모델의 성능은 88.8%가 나왔고, 제안 방식을 적용한 모델의 성능은 92.5%라는 높은 정확도에 도달하며 3.7%p의 성능 향상을 보여주었다. 또한 몸통 발차기 데이터셋에 대해서도 파라미터를 추가하기 이전에는 확장 합성곱을 적용한 평균 풀링 방식을 적용했을 때의 성능은 87.3%가 나왔으며, 제안 방식을 적용했을 때의 성능은 88.7%로 1.4%p의 성능 향상을 보여주었다.

다음은 VideoFocalNet 기본 모델과 선형 결합을 통해 노이즈를 추가한 모델, 모듈레이션 파라미터를 추가한 모델을 비교한 결과이다. 몸통 데이터에 대해서는 선형 결합 방식을 적용 시 성능 향상이 없었으나, 머리 발차기 데이터에 대해서는 성능 향상이 보임을 확인하였다.

머리 발차기 데이터셋에 대한 테스트 결과, 노이즈를 추가한 모델의 정확도가 91%로, 모듈레이션 파라미터를 추가한 모델의 정확도 92.5%에 비하여 낮은 성능을 보였으나, 기존 모델의 성능인 88.8%보다 높은 정확도를 보여주었다.

마지막으로 머리 발차기 데이터셋에 대하여 관련 연구에서 소개한 Slowfast, Mvit, Videofocalnet과 본 논문에서 제안한 각 프레임에 다음 프레임 정보를 선형 결합하는 노이즈 추가 모델(ours-n), 공간적 모듈레이션 정보와 시간적 모

표 1. 기존 모델과의 성능 비교
Table 1. Comparison of model performance

	Precision	Recall	F1-score	Accuracy
Slowfast	70.29%	75.89%	70.70%	70.15%
Mvit	77.38%	78.42%	75.52%	76.12%
VideoFocalNet (base)	87.36%	89.88%	88.34%	88.80%
VideoFocalNet (ours-n)	90.31%	90.10%	90.15%	91.04%
VideoFocalNet (ours-m)	92.05%	91.14%	91.43%	92.53%

표 2. VideoFocalNet(base)와 VideoFocalNet(ours-m) 모델 성능 비교
Table 2. Comparison of VideoFocalNet(base) and VideoFocalNet(ours-m) model performace

	VideoFocalNet(base)			VideoFocalNet(ours-m)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Normal Kick	81.81%	96.42%	88.52%	86.20%	89.28%	87.71%
An Chagi (Inward Crescent Kick)	96.55%	87.50%	91.80%	92.64%	98.43%	95.45%
Hook Kick	83.72%	85.71%	84.70%	97.29%	85.71%	91.13%

듈레이션 정보를 적절히 조절하는 모듈레이션 파라미터 추가 모델(ours-m)의 정확도 성능 비교를 수행했다. 합성곱 방식을 사용하는 Slowfast와 어텐션 방식을 사용하는 Mvit의 성능이 낮았고, 두 개의 모델에 비해 VideoFocalNet의 성능이 좋았다. 가장 성능이 좋은 모델은 본 논문에서 제안하는 모듈레이션 가중치를 추가한 기법을 적용한 VideoFocalNet 모델이었다.

다음으로는 기존 모델인 VideoFocalNet(base)과 가장 성능이 높은 머리 발차기 데이터셋을 이용한 VideoFocalNet (ours-m)과의 성능 비교를 통해 본 논문에서 제안하는 모델의 개선점을 명확히 확인할 수 있었다. 변칙 발차기에 대한 F1-점수가 전체적으로 향상되었고 정상 발차기의 재현율이 감소하고 정밀도가 증가하였다. 이를 통해 이상 발차기를 정상 발차기로 판단하는 잘못 판단하는 비율이 감소함을 확인할 수 있었다.

Ⅴ. 논 의

본 연구에서는 초점 변조 기반 동작 인식 모델에 시공간 가중치 조절 및 프레임 노이즈 결합 전략을 도입하여, 태권도 겨루기 상황에서의 변칙 동작 탐지 성능을 효과적으로 향상시켰다. 기존의 트랜스포머 기반 모델들은 모든 패치에 동일한 가중치를 적용하여 시공간 정보를 동일한 방식으로 처리하였기 때문에 미세한 동작 차이를 구분하는 데 한계를 보였다. 또한 기존 VideoFocalNet 구조는 초점 변조를 통해 시공간 정보를 통합적으로 처리할 수 있었으나, 시간 정보와 공간 정보의 상대적 중요도를 조정할 수 없는 한계가 있었다. 본 연구에서는 이러한 한계를 보완하기 위해 가변 모듈레이션 파라미터를 도입하여 시공간 가중치를 독립적으로 조절할 수 있는 구조를 설계하였다. 이를 통해 시공간 정보에서 미세한 차이를 보이는 동작에 대해 중요한 부분을 정교하게 파악하여 미세 동작 구분에서 뛰어난 성능을 보여주었다. 또한, 노이즈 주입 방식은 프레임 선형 결합을 통해 배경 정보에 대한 집중도를 낮추고 주요 객체의 동작 변화에 대한 모델의 민감도를 향상시켜 다양한 배경과 복잡한 환경에서도 실시간 동작 판별의 정확도를 확보할 수 있는 기초 기술로 활용될 수 있다. 또한 제안한 모델은 기존 트랜스포머 기반 모델과 비교하여 계산 복잡도 측면에서도 우수한 성능을 보였다. 기존 self-attention 방식은 $O(N^2D)$의 연산량을 요구하지만, 차원 합성곱을 이용하는 focal modulation을 통해 $O(ND)$의 선형 복잡도로 연산 효율성을 확보하였다. 하지만 비교적 정제된 환경에서 수집된 데이터셋을 기반으로 실험을 진행하였기 때문에, 다양한 조명 조건, 카메라 움직임, 불특정 다수 객체가 존재하는 실제 경기장에서의 일반화 성능은 향후 추가 실험이 필요하다.

본 연구는 단순한 모델 성능 향상을 넘어서, 실제 태권도 겨루기 판정 시스템에 적용 가능한 실용적 모델 구조를 제안했다는 점에서 높은 의의가 있다. 경계 상자 집계 알고리즘을 통해 선수 외 객체를 자동으로 제거함으로써 실제 경기 환경에서 불필요한 정보의 영향을 줄이고 인식 정확도를 높였다. 본 연구에서 제안한 알고리즘은 수작업 라벨링 없이도 효과적인 객체 크롭이 가능하므로, 시간과 비용을 절감할 수 있다. 이는 데이터 수집 및 처리 부담을 줄이고, 다양한 동작 인식 연구에서 생산성과 효율성을 향상하는데 기여할 수 있다.

Ⅵ. 결 론

본 연구에서는 태권도 겨루기 경기의 공정성 강화를 위해 태권도 비디오 데이터셋을 구축하고, 이를 활용하여 정상 발차기와 변칙 발차기를 효과적으로 구분할 수 있는 동작 인식 모델을 제안하였다. 미세한 시공간 정보를 포착하기 위해 노이즈 주입 방식과 가변 모듈레이션 파라미터를 이용한 전략을 통해 기존 모델 대비 성능이 향상됨을 확인하였고, 핵심적인 동작 정보를 더욱 효과적으로 학습할 수 있는 실시간 태권도 겨루기 경기에 맞는 객체 탐지 알고리즘을 설계하여 실용성 및 효율성을 확보하였다. 특히 가변 모듈레이션 파라미터를 적용한 모델은 변칙 발차기를 정상 발차기로 판단하는 오탐 비율을 감소시켜 실제 경기에서 부정 득점 방지에 더 효과적임을 입증하였다.

이러한 연구 결과는 정확한 시공간적 정보의 해석이 필수적인 다른 동작 인식 문제에도 적용될 가능성을 보여준다. 향후 연구에서는 본 모델을 다양한 동작 인식 비디오

데이터셋에 적용하여 일반화 성능을 분석하고, 데이터셋의 특성에 따라 최적의 프레임 결합 및 노이즈 부여 전략을 검증할 계획이다. 또한, 다양한 스포츠 및 행동 인식 데이터셋과의 비교 실험을 통해 모델의 확장 가능성을 평가하고, 동작 분류의 정밀도를 더욱 향상시킬 수 있는 방안을 모색할 예정이다.

태권도 겨루기는 전자 호구 및 영상 판독 시스템이 도입되었음에도 여전히 공정성 논란이 지속되고 있다. 특히 변칙 기술을 이용한 부정 득점 가능성이 남아있어, 더욱 정밀한 동작 분석 기술이 요구된다. 본 연구는 단순히 정확도를 개선하는 것에서 나아가 비디오 기반 동작 인식에서의 시공간적 정보 활용의 중요성을 재확인하고, 태권도 겨루기 경기에 최적화된 객체 탐지 및 동작 인식 모델 설계의 필요성을 제시한다는 점에서 의의가 있다. 또한 전자 호구 및 영상 판독 시스템의 한계를 보완하는 새로운 기술적 대안을 제시함으로써, 스포츠 판정 기술의 신뢰성을 향상시키는 데 기여할 수 있을 것으로 기대한다.

참 고 문 헌 (References)

[1] D. H. Kim, A comparative analysis of competition before and after the adoption of electronic truck protector for Taekwondo, Master's Thesis of Myongji University, Seoul, Korea, 2012 Retrieved from http://dcollection.mju.ac.kr/jsp/common/DcLoOrgPer.jsp?sItemId=000000058993

[2] World Taekwondo News, https://www.w-taekwondo.com/news/articleView.html?idxno=1486 (accessed Feb. 18, 2025)

[3] Takwondo Broadcasting System, http://www.itbstv.co.kr/news/articleView.html?idxno=2794 (accessed Feb. 18, 2025)

[4] World Taekwondo, COMPETITION RULES & INTERPRETATION, 2017.

[5] Taekwondo News, http://www.tkdnews.com/news/articleView.html?idxno=51641 (accessed Feb. 18, 2025)

[6] H.-S. Yang, J.-W. Yoon, and J.-Y. Kim, "Taekwondo International Competition Rules algorithm," Journal of Martial Arts, Vol.16, No.4, pp.99-118, 2022.
doi: https://doi.org/10.51223/kosoma.2022.11.16.4.99-118

[7] Dispatch, https://newsfeed.dispatch.co.kr/2156456 (accessed Feb. 18, 2025)

[8] Taekwondo News, http://www.tkdnews.com/news/articleView.html?idxno=6720 (accessed Feb. 18, 2025)

[9] H.-K. Lee, Biomechanical Comparison between Modified Kick and Roundhouse Kick, Master's Thesis of Chonnam National University, Kwangju, Korea, 2020. Retrieved from http://www.dcollection.net/handler/jnu/000000061857
doi: https://doi.org/10.23173/jnu.000000061857.24010.0000403

[10] Behind Sciences, https://behindsciences.kaist.ac.kr/2020/09/22/%EB%88%88%EC%B9%98-%EB%B3%B4%EB%A9%B0-%ED%95%98%EC%9D%B4%ED%82%A5-%EC%A0%84%EC%9E%90%ED%98%B8%EA%B5%AC%EB%8A%94-%ED%83%9C%EA%B6%8C%EB%8F%84%EB%A5%BC-%EB%A7%9D%EC%B3%90-%EB%86%93%EC%95%98%EB%8A%94/ (accessed Feb. 18, 2025)

[11] D.-B. Cho, H.-Y. Lee, and S.-S. Kang, "A Study on the Detection of Anomalous Kicks in Taekwondo games by using LSTM," Proceedings of Korea Information Processing Society, Vol.27, No.2, pp.1025-1027, 2020.

[12] A. Graves and A. Graves, "Long short-term memory," Supervised sequence labelling with recurrent neural networks, pp. 37-45, 2012 Springer.

[13] D.-B. Cho, H.-Y. Lee, W.-J. Lee, and S.-S. Kang, "Machine Learning based Detection of Anomalous Kicks in Taekwondo Match," Proceedings of Korean Institute of Information Scientists and Engineers, pp. 798-800, 2021.

[14] L.E. Peterson, "K-nearest neighbor," Scholarpedia, Vol.4, No.2, 2009.

[15] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," IEEE Intelligent Systems and their Applications, Vol.13, No.4, pp.18-28, 1998.
doi: https://doi.org/10.1109/5254.708428

[16] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 580-587.
doi: https://doi.org/10.1109/CVPR.2014.81

[17] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp.779-788.
doi: https://doi.org/10.1109/CVPR.2016.91

[18] C. S. Anumol, "Advancements in CNN Architectures for Computer Vision: A Comprehensive Review," 2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS), Kanjirapally, India, 2023, pp.1-7.
doi: https://doi.org/10.1109/AICERA/ICIS59538.2023.10420413

[19] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp.1725-1732.
doi: https://doi.org/10.1109/CVPR.2014.223

[20] J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 2625-2634.
doi: https://doi.org/10.1109/CVPR.2015.7298878

[21] C. Feichtenhofer, H. Fan, J. Malik and K. He, "SlowFast Networks for Video Recognition," 2019 IEEE/CVF International Conference on

Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 6201-6210.
doi: https://doi.org/10.1109/ICCV.2019.00630
[22] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić and C. Schmid, "ViViT: A Video Vision Transformer," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 6816-6826.
doi: https://doi.org/10.1109/ICCV48922.2021.00676
[23] H. Fan et al., "Multiscale Vision Transformers," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 6804-6815.
doi: https://doi.org/10.1109/ICCV48922.2021.00675
[24] S. T. Wasim, M. U. Khattak, M. Naseer, S. Khan, M. Shah and F. S. Khan, "Video-FocalNets: Spatio-Temporal Focal Modulation for Video Action Recognition," 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023, pp.13732-13743.
doi: https://doi.org/10.1109/ICCV51070.2023.01267

저 자 소 개

이 하 랑

- 2023년 2월 : 서울호서대학교 사이버해킹보안과 졸업(학사)
- 2025년 2월 : 건국대학교 스마트 ICT융합공학과 졸업(석사)
- ORCID : https://orcid.org/0009-0002-6344-9000
- 주관심분야 : 영상처리, 인공지능, 컴퓨터 비전

윤 경 로

- 1987년 2월 : 연세대학교 전자전산기공학과 졸업(학사)
- 1989년 12월 : University of Michigan, Ann Arbor, 전기전산기공학과 졸업(석사)
- 1999년 5월 : Syracuse University, 전산과학과 졸업(박사)
- 1999년 6월 ~ 2003년 8월 : LG전자기술원 책임연구원/그룹장
- 2003년 9월 ~ 현재 : 건국대학교 컴퓨터공학부 교수
- 2017년 10월 ~ 현재 : 국립전파연구원 멀티미디어부호화 전문위원회 (JTC1 SC29K) 대표전문위원
- 2019년 10월 ~ 현재 : IEEE SA 2888 WG 의장
- 2024년 10월 ~ 현재 : IEEE SA 디지털콘텐츠기술 표준화 위원회 의장
- ORCID : https://orcid.org/0000-0002-1153-4038
- 주관심분야 : 스마트미디어시스템, 멀티미디어검색, 영상처리, 컴퓨터비전, 멀티미디어/메타데이터 처리

심 영 균

- 2020년 2월 : 단국대학교 일반대학원 체육학과 졸업(박사)
- 2022년 1월 ~ 현재 : 우석대학교 산학협력단 체육복지융합연구소 부소장
- 2016년 6월 ~ 현재 : World Taekwondo Federation International Kyorugi Referee
- 2023년 12월 ~ 현재 : World Taekwondo Federation International Poosae Referee
- 2023년 6월 ~ 현재 : World Taekwondo Federation Kyorugi Educator
- 2024년 8월 : Paris 2024 Paralympic Games Referee(Taekwondo)
- ORCID : https://orcid.org/0000-0001-8249-9121
- 주관심분야 : 태권도, 경기분석, 측정평가