

# VONV: VSR 품질 보장과 열 제한 영향 완화를 위한 NR-VMAF 기반 성능 최적화

박혜진\*, 하란°

## VONV: NR-VMAF Based Optimization for VSR Quality Assurance and Thermal Throttling Mitigation

Hye-jin Park\*, Rhan Ha°

요약

On-Device VSR(Video Super Resolution)은 높은 연산 부하로 인해 발열과 성능 저하 문제가 발생한다. 이를 해결하기 위한 동적 VSR에서는 단순한 성능 유지뿐 아니라 영상 품질까지 함께 고려한 균형 잡힌 접근이 필요하다. 본 연구에서는 검증된 품질 평가 지표인 VMAF(Video Multimethod Assessment Fusion)를 예측하는 NR(No-Reference)-VMAF를 활용하여 프레임별로 DNN 기반 SR과 bicubic 보간을 선택적으로 적용함으로써 열 제한의 영향을 완화하며 일정한 품질을 유지하는 동적 VSR 기법인 VONV(VSR Optimization with NR-VMAF)를 제안한다. VONV에서 예측한 NR-VMAF는 실제 FR(Full-Reference)-VMAF와 오차가 2점 이내로 높은 정확도를 보이며 품질 예측의 신뢰성이 검증되었다. 또한 성능 차이가 명확한 두 기기에서 VONV의 성능 최적화 효과를 확인하였을 때 두 기기 모두에서 열 제한 발생 시점을 약 3배 지연시키고 전체 추론 시간을 최소 1.4배에서 최대 4배까지 단축하였다. 이는 VONV가 VMAF 예측을 활용하여 모바일 환경에서 품질과 성능의 최적화를 달성하는 효과적인 VSR 개선 방법임을 의미한다.

**키워드** : 초해상화, 발열 관리, 최적화, 온디바이스 AI, 모바일 시스템

**Key Words** : Super Resolution, Thermal Management, Optimization, On-Device AI, Mobile System

### ABSTRACT

On-Device VSR(Video Super Resolution) is challenged by the high computational load of Deep Neural Networks(DNNs), causing heat and performance degradation. Adaptive VSR offers a solution by optimizing performance, but it must also take a balanced approach that considers not only maintaining performance but also ensuring visual quality. This paper proposes VONV(VSR Optimization with NR-VMAF), an adaptive VSR method that uses NR(No-Reference)-VMAF to selectively apply DNN-based SR or bicubic interpolation per frame, in order to mitigate thermal throttling while maintaining consistent video quality. NR-VMAF demonstrates high accuracy, staying within 2 points of FR-VMAF, validating its reliability. Tests on two devices with distinct performance profiles showed VONV extends the thermal throttling threshold threefold and reduces overall inference time by 1.4 to 4 times. These results demonstrate that VONV, by using NR-VMAF, effectively optimizes both quality and performance in mobile environments.

\* First Author : Hongik University Department of Computer Engineering, hyejin2371@gmail.com 학생회원

° Corresponding Author : Hongik University Department of Computer Engineering, rhanha@hongik.ac.kr, 정회원

논문번호 : 202502-026-E-RN.R1, Received February 3, 2025; Revised April 6, 2025; Accepted April 11, 2025

## I. 서 론

최근 실시간 비디오 스트리밍은 엔터테인먼트, 교육 콘텐츠, 소셜 미디어 등 다양한 산업 분야에서 중요한 역할을 하고 있다. 사용자 기대를 만족시키기 위해 네트워크 인프라 개발과 비디오 전송 기술의 발전이 이루어졌으나 비디오 스트리밍의 품질은 네트워크 대역폭에 크게 의존한다. 딥러닝 기술과 클라이언트 기기의 발전이 이루어지며 고품질 비디오 스트리밍을 제공하기 위한 방법으로써 VSR(Video Super Resolution)이 주목받고 있다<sup>[1]</sup>. DNN(Deep Neural Network)을 사용해 저해상도 비디오를 고해상도로 재생성하는 VSR은 엣지 디바이스에서 비디오 스트리밍에 요구되는 네트워크 대역폭을 크게 낮추고 전송 지연을 최소화한다. 하지만 공간적 제약으로 사용 가능한 냉각 시스템이 제한적인 모바일 기기에서의 DNN 연산은 높은 연산 부하로 쉽게 기기의 과열을 유발한다.

발열은 엣지 디바이스의 성능 변동성을 유발하는 주요 요인으로, 사용 중인 기기의 성능에 직접적이고 즉각적인 영향을 미친다. 대부분의 모바일 기기는 발열 해소를 위해 기기의 성능과 온도를 동적으로 관리하는 DTM(Dynamic Thermal Management)을 사용한다. DTM은 온도를 모니터링하고 기기의 전력 소비, 프로세서의 속도를 조정하는 등의 방식으로 성능과 발열 제어 간의 균형을 유지한다. DVFS(Dynamic Voltage and Frequency Scaling)는 보편적으로 사용되는 DTM 기법으로, 모바일 기기의 온도가 임계값까지 상승하면 프로세서의 클럭을 낮춰 전력 소모와 발열을 억제한다. DVFS는 모바일에서 효과적인 발열 관리 방법이지만 열 제한 발생 후 GPU 클럭이 회복되기까지 오랜 시간이 소요되며 성능이 감소하는 문제가 있다. 또한 주로 운영체제에서 DVFS를 관리하기 때문에 애플리케이션의 특성을 고려하지 않고 프로세서의 클럭 속도를 낮춰 불필요한 성능 저하를 유발할 수 있다. 따라서 애플리케이션의 요구사항에 따라 연산 강도를 적절히 조절해야 제한된 자원 내에서 최적의 성능을 제공할 수 있다.

On-Device VSR은 네트워크 대역폭 요구를 크게 줄이지만 강한 연산 부하로 인한 발열과 운영체제 중심의 발열 제어 정책은 사용자 만족도 저하를 유발한다. 이러한 문제를 해소하기 위해 기기의 상태, 조건에 따라 성능을 최적화하는 동적 VSR이 활발히 연구되고 있다. 특히 모바일 기기에서는 발열의 영향이 크기 때문에 이를 고려한 VSR 전략이 필수적이다. VSR 비디오의 품질 역시 사용자 경험에 중요한 영향을 미치므로 품질을 보장하면서 열 제한의 영향을 최소화하기 위한 균형잡

힌 접근법이 요구된다.

VMAF(Video Multimethod Assessment Fusion)<sup>[2]</sup>는 여러 평가 지표를 종합적으로 반영하는 새로운 비디오 평가 방법으로 전통적인 방법보다 인간의 지각을 더 정확하게 표현하나 원본 비디오를 참조하는 FR(Full-Reference) 방식에 의존한다. 이를 해결하기 위해 제시된 NR(No-Reference)-VMAF는 비디오의 구조적 특징이나 DNN을 사용하여 손상된 영상만으로 VMAF 점수를 예측한다. NR-VMAF는 전통적인 품질 평가 지표보다 인간의 지각을 더 정확하게 반영하며 참조 영상을 구하기 어려운 스트리밍 서비스, 모바일 환경의 비디오 품질 평가에 유용하게 사용된다<sup>[3]</sup>. 본 연구에서는 NR-VMAF 기반의 동적 VSR을 통해 발열을 최소화하며 양호한 품질을 보장하는 VONV(VSR Optimization with NR-VMAF)를 제안한다.

VONV는 프레임마다 DNN 기반 SR과 bicubic 보간을 선택적으로 적용하여 연산을 최적화함으로써 열 제한 도달 시점을 늦추고 열 제한 발생 빈도를 줄인다. 결과적으로 열 제한으로 인한 성능 저하를 완화하여 단일 SR 수행 대비 비디오 스트리밍의 효율성과 안전성을 높인다. SR 방법 선택 전략은 유효하게 추정된 NR-VMAF를 기반으로 이루어져 연산량을 줄이는 동시에 영상 품질을 보장할 수 있도록 설계되었다. 2장에서는 발열 완화를 위한 동적 VSR 전략과 NR-VMAF에 관한 선행 연구를 분석하고 3장에서 VONV의 전체 프로세스, 사용된 VSR 모델, NR-VMAF 모델을 설명한다. 4장에서는 NR-VMAF 기반 품질 평가의 신뢰성을 확인하고 VONV 적용 전후의 온도 변화와 추론 시간을 비교하여 VONV의 효율성과 열 제한 영향 완화 효과를 검증한다.

## II. 관련 연구

VSR은 효과적으로 대역폭 제한과 네트워크 부담을 줄이는 전략이지만 이때 사용되는 DNN 연산은 기기를 쉽게 과열시킨다. 모바일 기기가 과열되면 프로세서의 클럭 속도가 낮아지고 온도가 회복될 때까지 성능이 저하되므로 안정적인 품질을 제공하기 어렵다. 따라서 열 제한의 영향을 줄이고 안정적인 품질을 제공하기 위해 조건에 맞춰 VSR을 동적으로 수행하는 방법이 주목받고 있다. OmniLive<sup>[4]</sup>는 multi-exit DNN 구조를 활용하여 열 제한으로 인해 변동하는 추론 시간을 예측하고 그에 따라 연산 단계를 조절하여 안정적인 360° 비디오 실시간 스트리밍을 제공했다. ATM<sup>[5]</sup>은 기기의 실시간 온도에 따라 SR 모델의 연산 단계를 조절함으로써

On-Device VSR에서 기기의 열 제한 발생을 효과적으로 지연시키고 추론 시간을 단축하였다.

위 연구를 비롯한 On-Device VSR 최적화 연구에서는 전통적인 VQA(Video Quality Assessment)인 PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity Index Measure)이 성능 평가에 주로 사용된다. 이러한 지표들은 이미지의 구조적 특성을 수학적으로 계산하여 실제 사용자의 지각을 제대로 반영하지 못한다고 지적되었다. 구체적으로 PSNR은 픽셀 간 차이를 통해 품질을 측정하지만 모든 픽셀 간 차이를 동일하게 취급하기 때문에 인간의 시각적 인지 특성을 제대로 반영하지 못한다. SSIM은 PSNR의 한계를 보완하기 위해 밝기, 명암 대비, 구조적 유사성의 3가지 측면에서 품질을 측정하여 PSNR보다 유의미한 결과를 제공하나, 여전히 인간의 복잡한 왜곡 요소를 완전히 반영하지 못한다. 결정적으로 두 지표는 인간의 주관적 평가인 DMOS(Difference Mean Opinion Score)와 정확한 비례 관계를 보이지 않으며<sup>3)</sup>, 특히 두 지표의 품질 차이가 DMOS와 비교하여 일정하지 않다. 이는 두 지표가 소폭 변할 때도 실제 시각적 품질에 큰 차이가 느껴지거나 반대로 큰 수치 차이에도 체감 품질은 미미할 수 있음을 의미한다<sup>3)</sup>.

VMAF는 이러한 한계를 극복하기 위해 제안된 평가 지표로, 인간이 인지하는 디테일 손실, 움직임의 왜곡을 반영하며 VIF(Visual Information Fidelity)<sup>6)</sup>, DLM(Diagonal Laplacian Measure)<sup>7)</sup>, Motion을 기본 metric으로 사용한다. VIF는 영상의 시각적 정보가 시간 변화 속에서 보존되는 정도를 측정하고, DLM은 윤곽선과 질감의 변화를 정량화하여 영상의 세밀한 구조를 측정하는 지표이다. Motion은 관련된 프레임 사이에 픽셀 차이를 포착해 움직임이 영상 품질에 미치는 영향을 반영한다. VMAF는 FR(Full-Reference)-VQA로 변형된 영상과 원본 영상을 프레임 단위로 비교하기 때문에 참조 영상이 없거나 연산 자원이 부족한 환경에서는 사용이 어렵다. 이에 따라 게임, 모바일 스트리밍 등의 분야에서 VMAF를 활용하기 위해 원본 자료 없이 점수를 예측하는 NR-VMAF가 연구되었다<sup>4)</sup>.

VMAF 예측을 위한 NR-methods는 특징 기반 접근법과 DNN 기반 접근법 두 가지로 나뉜다. 특징 기반 접근은 주로 NSS(Natural Scene Statistics) 기반으로, 원본 영상의 규칙적인 특성이 왜곡된 영상에서 얼마나 변형되었는지 측정하여 품질을 추정한다. DNN 기반의 접근법은 주로 CNN을 이용해 공간적 특징을 추출하고 RNN으로 시간적 정보를 추적하는 것이 일반적이다. 기존 NR-VMAF 연구들의 데이터셋 부족 문제를 해소

하기 위해 패치 기반 처리 방식을 사용하고, 주관적 평가 점수인 DMOS를 직접 예측하는 대신 VMAF 지표를 예측하는 연구도 시도되었다<sup>4)</sup>.

기존의 NR-VMAF 연구는 VMAF를 예측하는 새로운 방법 개발에 집중하고 이를 비디오 품질 평가에 활용하는 데에 그쳤다. VONV는 NR-VMAF를 사전 품질 예측과 동적 VSR 운영에 활용한다는 점에서 이전의 NR-VMAF 연구들과 차별화된다. 또한 발열 영향 완화가 목적인 기존 동적 VSR은 전통적 품질 지표로 비디오 품질을 평가하여 인간의 실제 지각을 충분히 반영하지 못하는 한계가 있다. VONV는 동적 VSR을 통해 열 제한의 영향을 줄이면서 VMAF를 예측하여 사용자의 시각적 품질을 고려한다. 이로써 QoE를 보장하며 On-Device VSR의 효율성과 안정성을 향상시킨다.

### III. 본 론

VONV는 효과적인 비디오 품질 평가 지표인 VMAF 예측을 바탕으로 프레임의 특징에 따라 동적 SR을 수행한다. NR-VMAF 기반으로 프레임에 최적화된 SR 방법을 선택하여 높은 품질을 유지하면서 연산 부하를 최적화한다. 이를 통해 단일 SR 방식보다 열 제한 영향을 완화하며 추론 시간을 단축하는 효과를 함께 제공한다.

#### 3.1 VONV의 구조 및 SR 모델 개요

그림 1은 VONV의 전체 프로세스를 나타낸다. 스트리밍이 시작되면 VONV의 NR-VMAF 모델은 서버에서 저해상도 프레임을 입력받아 bicubic 기반 업샘플링, DNN 기반 SR 수행 시의 VMAF를 예측한다. 일반적으로 VMAF는 0부터 100까지의 값으로, 70점부터 양호한 품질, 80점 이상부터 우수한 품질로 해석된다<sup>8)</sup>. VONV는 우수한 품질을 유지하며 발열을 완화하기 위해 서버에서 각 프레임의 처리 방법 별 VMAF를 예측하고 80점 이상의 품질을 보장하는 방법을 선택한다. 이로써 복잡한 처리 방법이 필요하지 않은 프레임에는 상대적으로 계산 비용이 적은 bicubic 기반 업샘플링을 사용하고 더 많은 연산이 요구되는 프레임은 DNN 기반 SR을 사용해 연산의 효율성을 높인다. 이후 서버는 예측한 VMAF 기반으로 결정된 처리 방법과 저해상도 프레임을 기기에 전송한다. 사용자 기기는 서버로부터 저해상도 프레임과 처리 방법을 수신한 후 프레임별로 해당 방법을 적용하여 고해상도 비디오를 생성한다. 서버는 VMAF 예측과 비디오 처리 방식 선택을 담당하고 모바일 기기는 프레임 변환만 수행하여 연산 부담을 줄

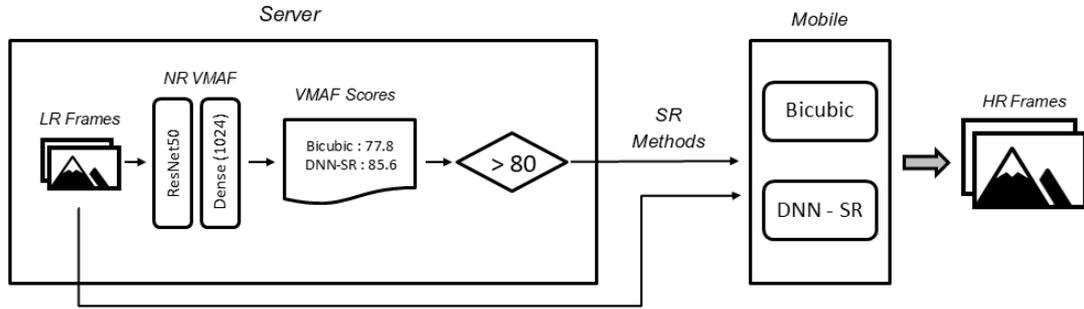


그림 1. VONV의 전체 프로세스  
Fig. 1. Entire process of VONV

이다. 또한 기기는 bicubic 기반 업샘플링을 함께 사용하여 DNN 기반 SR을 단독으로 수행할 때보다 연산 부하를 분산시키고 발열 증가 속도를 늦춘다.

본 연구에서 사용된 VSR 모델 SeeMore<sup>[9]</sup>는 전역적 특징과 지역적 특징을 모두 고려하여 효율적인 SR을 수행한다. 먼저 채널별 픽셀 관계와 공간적 특징을 병렬로 추출한 뒤 저차원에서 결합하고 이를 striped-depth-wise 컨볼루션 레이어를 이용하여 지역적 공간 특징을 보완한다. Striped-depth-wise 컨볼루션은 SeeMore에서 사용된 용어로,  $k \times k$  크기의 커널을  $1 \times k$  와  $k \times 1$ 로 나누어 차례로 컨볼루션을 수행하는 방식이다<sup>[11]</sup>. 저차원 특징 결합은 1차원부터 n차원까지 총 n개의 모듈에서 병렬로 이루어지고 실제 추론 시에는 이중 가장 적합한 차원의 모듈이 활성화된다. 훈련 단계에서 각 모듈이 입력 데이터를 각 차원으로 압축하여 개별적으로 연산을 진행하면서 입력 데이터의 특성에 가장 적합한 차원을 선택하도록 학습된다. 이로써 입력 이미지의 특성을 반영할 수 있는 차원을 동적으로 선택하여 연산 효율성을 향상시킨다.

SeeMore는 효율적이고 높은 성능을 가진 VSR 모델이지만 모바일 환경에서는 연산 부담이 커 본 연구에서는 모델을 경량화하여 사용한다. 수정된 VSR 모델은 입력 데이터의 임베딩 차원을 36에서 4로, 저차원 특징 결합의 반복 횟수는 6회에서 1회로 축소되었으며 지역적 정보 보완 시 사용되는 컨볼루션 커널의 크기도 11에서 5로 감소하였다. 학습은 학습률 0.01, 배치 크기 64에서 Adam optimizer, REDS 데이터셋<sup>[10]</sup>으로 진행되었다. 이로써 프레임의 평균 품질은 VMAF 91에서 VMAF 85로 감소했지만 평균 VMAF 77.5인 bicubic 기반 업샘플링 방법에 비해 향상된 품질을 보이며 추론 시간을 평균 800 ms에서 56 ms로 크게 단축하였다. 이로써 원본 모델의 추론 시간과 품질 간 균형을 조정하여 모바일 환경에 적합한 실용적인 성능을 달성하였다.

### 3.2 NR-VMAF for VONV

VONV의 NR-VMAF 모델은 저해상도 프레임을 입력받아 각 처리 방식에 따른 VMAF를 예측한다. 본 연구에서는 NR-VMAF 모델 설계를 위해 몇 가지 구조를 실험적으로 비교하고 적절한 모델 구조를 선정하였다. 최적의 CNN backbone을 탐색하기 위해 ResNet<sup>[11]</sup>, EfficientNet<sup>[12]</sup>에 2개의 출력 브랜치를 더해 예측 정확도를 비교하였다(표 1). 다음으로 선택된 backbone에 최종 출력용 FC(Fully Connected) 레이어 구조를 비교하여 최종 모델 구조를 결정하였다(표 2). NR-VMAF 예측은 연산 자원이 충분한 서버에서 이루어지므로 정확도를 우선으로 CNN backbone을 선정하였다.

실험은 TensorFlow 기반으로 서버에서 수행하였고, 720p 해상도의 REDS 데이터셋을 학습(80%)과 검증(20%) 용도로 나누어 사용하였다. 사전에 계산한 FR-VMAF 점수를 저해상도 프레임에 라벨링하여 훈련 데이터로 활용하였다. 예측값의 변동성이 크지 않아 이상치의 영향이 적으므로 훈련 및 검증 오차는 MAE(Mean Absolute Error)를 사용하였으며 검증 오차를 기준으로 정확도를 비교하였다.

먼저 CNN backbone 비교를 위해 저해상도 프레임을 입력받아 두 가지 SR 방법의 VMAF 점수를 각각 예측한 결과, ResNet이 약 2점으로 가장 작은 오차를 보여 backbone으로 선정되었다(표 1). 이는 실제 FR-VMAF가 74점일 경우 NR-VMAF 모델이 약 72점

표 1. CNN backbone 별 VMAF 예측 오차  
Table 1. VMAF prediction error by CNN backbone

	ResNet	EfficientNetB3	EfficientNetB2
Bicubic Upsampling	2.79	3.79	8.27
DNN based SR	2.08	2.93	8.66

표 2. FC 레이어의 구조 별 NR-VMAF 예측 오차  
Table 2. NR-VMAF prediction error by FC layer structure

FC (Units)	Bicubic Upsampling	DNN based SR
1 Layer (1024 units)	2.03	1.71
2 Layers (1024, 512 units)	3.27	2.72
3 Layers (1024, 512, 256 units)	3.99	2.45

에서 76점 사이의 점수를 예측함을 의미한다. VMAF 기준에서 6점 차이는 시각적으로 인지 가능한 차이를 나타내는 JND(Just Noticeable Difference) 1점에 해당하므로<sup>[13]</sup>, 예측된 NR-VMAF는 FR-VMAF와 높은 일치도를 보이며 오차는 인지적으로 유의미한 차이를 발생시키지 않는 수준이다.

이후 backbone에 추가될 FC 레이어 구조 별 예측 오차를 비교하였다. 표 2는 앞서 정해진 CNN 백본에 각각 1개(1024 유닛), 2개(1024·512 유닛), 3개(1024·512·256 유닛)의 FC 레이어를 추가했을 때의 검증 오차를 나타낸다. 실험 결과, ResNet backbone에 1024개 유닛을 갖는 하나의 FC 레이어를 추가한 경우 가장 낮은 오차를 보여 최종 구조로 채택되었다. 최종 NR-VMAF 모델의 성능은 ‘IV. 1. NR-VMAF 기반 품질 보장 효과’에서 더 자세히 평가한다.

#### IV. 성능 평가

VONV의 효과를 확인하기 위해 같은 실험 환경에서 VONV 적용 시와 미적용 시의 VSR 성능 차이를 비교한다. VONV를 미적용한 경우 III-1 절에서 설명한 바와 같이 모바일 환경에 적합하도록 경량화된 SR 모델을 단독으로 사용하였다. 그리고 VONV를 적용한 경우에는 동일한 SR 모델을 사용하되 예측된 VMAF 점수를 기준으로 bicubic 보간을 병행하였다. 다양한 구조의 이미지를 처리하기 위해 REDS의 테스트 데이터셋, DIV2K<sup>[14]</sup>, Unsplash2k<sup>[15]</sup>를 증강시켜 약 8,000장의 데이터를 실험에 사용한다. 데이터셋은 회전, 색상 조정 등을 통해 증강하여 모델의 일반화 성능을 높였다. NR-VMAF 모델과 적응적 VSR의 효과를 검증하기 위해 다음과 같이 두 가지 실험을 진행하였다.

첫째, VMAF 예측을 통한 비디오 품질 보장 효과를 확인하기 위해 NR-VMAF 모델의 성능을 평가한다. 모바일 기기의 저장 공간 제약으로 인해 실시간으로 SR

된 프레임은 저장할 수 없으므로 FR-VMAF 계산을 위한 SR 프레임은 서버상에서 동일한 VSR 모델을 따로 수행하여 생성한 결과를 사용하였다. 둘째, VONV에서 적응적 VSR을 통한 열 제한 영향 완화 효과를 확인하기 위해 같은 조건에서 VSR 수행 시 VONV 적용 전후의 추론 시간, 기기의 온도, 열 제한 상태를 관찰한다. 적응적 VSR의 성능 평가는 성능 차이가 명확한 Galaxy Note 20, Galaxy S23 상에서 진행되었으며 기기에는 각각 Adreno GPU 650과 750이 탑재되어 있다. 기기의 온도와 열 제한 발생 여부는 Android Thermal API를 통해 측정되었다.

#### 4.1 NR-VMAF 기반 품질 보장 효과

먼저 검증된 화질 평가 지표인 FR-VMAF를 PSNR, SSIM과 비교하여 전통적인 VQA의 한계를 분석한다. 이상적인 경우 두 값은 VMAF와 동일한 방향으로 변하고 변화량의 크기 또한 유사해야 한다. 이를 검증하기 위해 bicubic 업스케일을 사용한 프레임과 DNN 기반 SR을 적용한 프레임의 PSNR, SSIM, VMAF 차이를 분석하였다. Bicubic 업스케일과 DNN 기반 SR 수행 시 PSNR, SSIM 차이와 VMAF 차이의 상관관계를 나타내는 그림 2에서 PSNR, SSIM은 VMAF의 변화에 대해 규칙적인 분포를 보이지 않으며 VMAF의 관계성이 부족함을 직관적으로 파악할 수 있다. 정량적인 분석을 위해 각 품질 평가 지표 차이의 PLCC(Pearson Linear Correlation Coefficient)를 계산하였다. PLCC는 -1에서 +1 사이의 값을 가지며 양수일 때는 두 변수 간의 상관도가 비례하고 음수일 때는 반비례함을 나타낸다. 값의 절댓값이 클수록 두 변수 간의 선형적 상관관계가 강하다고 해석된다. 그림 2에서 PSNR과 VMAF의 PLCC는 -0.39, SSIM과 VMAF의 PLCC는 -0.105로 나타나 PSNR SSIM으로 측정된 프레임의 품질

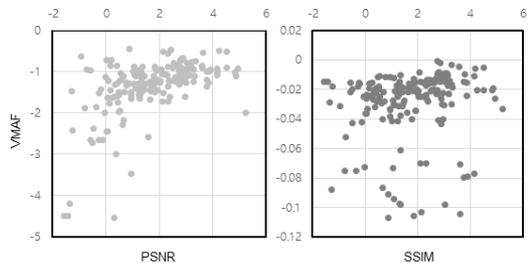


그림 2. Bicubic 업스케일, DNN 기반 SR 프레임의 PSNR, SSIM, VMAF 오차의 상관관계  
Fig. 2. Correlation between PSNR, SSIM differences and VMAF difference between bicubic upscaled and DNN-based SR frames

질 차이가 FR-VMAF와 매우 약한 상관관계를 보였다. 따라서 전통적인 두 지표가 화질을 정교하게 설명하기에 부족함을 알 수 있다.

한편, VONV에서 사전 품질 예측을 위한 NR-VMAF 모델의 예측 결과와 FR-VMAF를 비교한 결과(표 3), 두 값의 절대 오차는 약 2.6으로 시각적으로 품질 차이가 인지되는 기준인 6점보다 낮아 정확한 예측 성능을 보였다. NR-VMAF와 FR-VMAF의 PLCC는 bicubic 업샘플링과 DNN 기반 SR이 적용된 프레임 모두에서 0.8 이상으로 두 값의 선형적 상관관계가 명확함을 나타냈다(그림 3). 두 데이터 집합 간의 순위 관계를 -1부터 +1까지의 값으로 나타내는 SROCC(Spearman Rank Order Correlation Coefficient) 또한 모든 경우 0.8 이상으로, NR-VMAF와 FR-VMAF의 순위 상관관계가 명확하게 일치함을 보여준다.

두 평가 지표의 오차에 비해 상관관계가 매우 강하지 않은 이유는 두 점수의 분포에 기인한다. 그림 3처럼 두 점수가 70과 90 사이에 집중적으로 분포하면 분산이 작아져 PLCC에서는 데이터의 상대적 변화가 잘 드러나지 않는다. SROCC 역시 값의 차이가 크지 않으면 순위 관계가 뚜렷하게 나타나지 않아 상관관계가 약하게 표현된다. 그러나 오차가 6점 이하로 시각적 차이가 없어 NR-VMAF는 FR-VMAF를 대체할 수 있으며 참

표 3. FR-VMAF와 NR-VMAF의 오차, 상관 관계 비교  
Table 3. Comparison of error and correlation between FR-VMAF and NR-VMAF

	avg. MAE	PLCC	SROCC
Bicubic Upsampling	2.661	0.816	0.812
DNN based SR	2.608	0.823	0.813

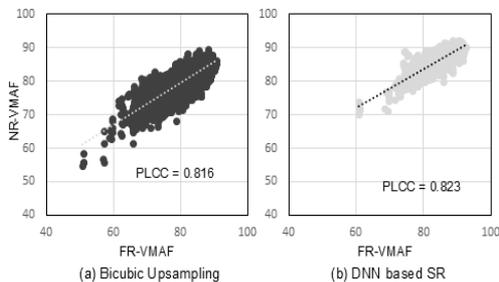


그림 3. Bicubic 업스케일, DNN 기반 SR 시 FR-VMAF와 NR-VMAF의 상관관계  
Fig. 3. Correlation between FR-VMAF and NR-VMAF for Bicubic Upscaling and DNN-based SR

조 영상 없이도 스트리밍 환경에서 신뢰도 높은 품질 예측이 가능함을 확인했다.

#### 4.2 발열 지연과 추론 시간 단축 효과

적응적 VSR을 통한 모바일 기기의 발열 지연 효과와 추론 시간 단축 효과를 분석하기 위해 약 15분간 서버와 기기에서 연속적으로 프레임을 처리했다. 먼저 기기의 온도 변화를 관찰한 결과, Galaxy Note 20는 VONV를 미적용한 경우 131초에 열 제한이 발생했다(그림 4). 발열이 일정 수준 이상 발생하면 시스템 자체의 DVFS 정책에 의해 GPU 클럭 속도가 급격히 낮아지며 온도를 점차 회복한다. 이에 따라 GPU 클럭이 크게 하락하며 평균 추론 시간이 1,094 ms에서 1,796 ms로 약 1.8배 증가한다. DVFS를 통해 온도가 정상화 되면 추론 속도가 함께 회복되지만 다시 기기가 과열되며 반복적으로 열 제한 상태에 빠진다.

반면 Galaxy Note 20에서 VONV를 적용한 경우 처음 열 제한이 415초에 발생하여 적용하지 않았을 때

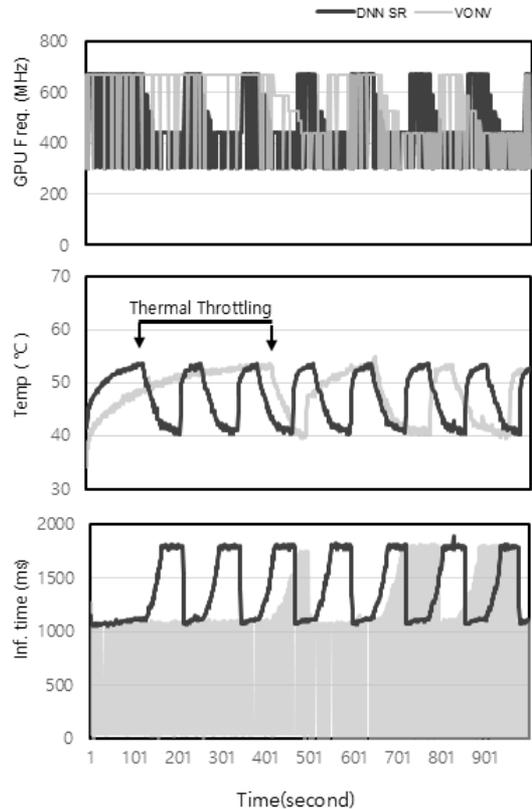


그림 4. Galaxy Note 20에서 VONV 적용 유무에 따른 추론 시간과 온도 변화  
Fig. 4. Inference time and temperature changes with and without VONV on Galaxy Note 20

다 열 제한 발생 시점을 약 3.16배 지연시키는 효과를 보였다. VONV는 첫 열 제한 발생 시점을 늦출 뿐 아니라 이후 열 제한이 재발하는 횟수 또한 6회에서 2회로 감소시켰다. 더불어 입력 데이터의 특징에 따라 연산 비용이 매우 적은 bicubic 업샘플링을 함께 사용함으로써 평균 추론 시간이 1,393 ms에서 348 ms로 크게 단축되었다.

그림 5는 Galaxy S23에서 VONV의 유무에 따른 발열 제어 효과와 추론 시간 변화를 측정한 그림이다. VONV를 적용하지 않은 경우 작업 초기에 고강도 연산을 집중적으로 수행하며 기기의 온도가 빠르게 상승하고 약 94초에서 열 제한이 발생한다. 반면 VONV 적용 시에는 연산 부하가 조절되어 온도 증가 속도가 감소하고 274초에서 열 제한이 발생하여 열 제한 발생 시점을 약 2.9배 지연시켰다.

Galaxy Note 20에 비해 Galaxy S23은 고성능 프로세서를 통하여 보다 안정적인 성능을 제공한다. 따라서 열 제한이 발생하더라도 GPU 클럭이 저하되지 않으며

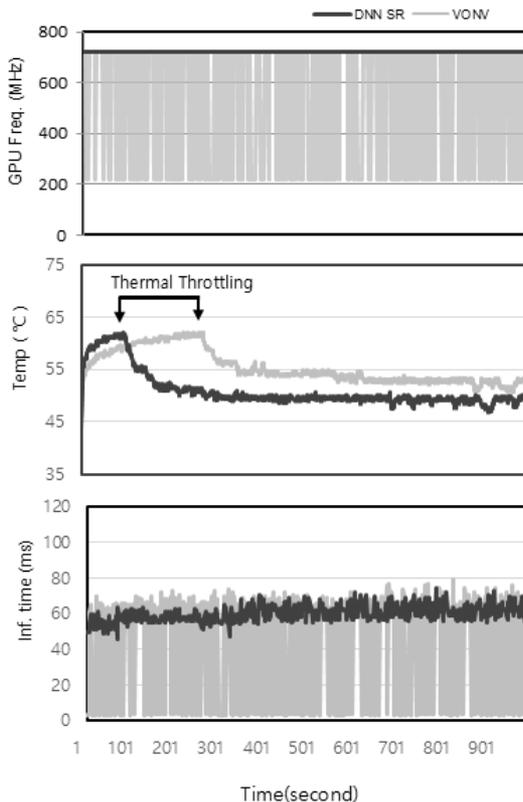


그림 5. Galaxy S23에서 VONV 적용 유무에 따른 추론 시간과 온도 변화  
 Fig. 5. Inference time and temperature changes with and without VONV on Galaxy S23

열 제한 전후의 추론 시간 변화가 미미하다. VONV를 적용한 경우 열 제한 발생 전후 추론 시간이 약 39 ms에서 42 ms로, VONV를 미적용한 경우에는 56 ms에서 59 ms로 소폭 증가했다. 실험에서 열 제한 발생 전후의 추론 시간 변화는 크지 않았으나 VONV를 적용하며 bicubic 업샘플링을 병행함으로써 프레임당 추론 시간이 평균 60 ms에서 42 ms로 크게 단축되었다.

결론적으로 VONV는 동적 VSR을 통해 열 제한 영향을 효과적으로 완화하였다. 발열 제어 성능이 약한 Galaxy Note 20에서는 bicubic 업샘플링과 DNN 기반 SR을 선택적으로 적용해 열 제한으로 인한 성능 저하를 완화하고 추론 시간을 크게 줄였으며, Galaxy S23에서도 열 제한 지연과 전반적인 추론 시간 감소 효과를 확인하였다.

### 4.3 성능 평가 요약

본 논문에서 제안한 VONV는 유효한 품질 평가 지표인 NR-VMAF로 프레임의 품질을 예측하고 이를 바탕으로 최적의 SR 방법을 선택적으로 수행한다. 인간의 지각을 정밀하게 반영하는 VMAF를 예측하여 기존 비디오 평가 지표보다 정교한 사전 예측을 수행하고 품질을 보장한다. NR-VMAF 기반 동적 VSR을 수행하여 VSR의 안정성과 효율성을 향상시키고 성능 최적화와 비디오 품질의 균형을 효과적으로 유지한다.

사전 품질 예측의 신뢰도를 평가하기 위해 예측한 NR-VMAF를 FR-VMAF와 비교한 결과, 2점 이내의 오차와 PLCC 0.81, SROCC 0.82의 높은 통계적 연관성과 정확도를 보여준다. 이는 NR-VMAF가 FR-VMAF와 유사한 표현력을 가지며, 품질 예측의 신뢰성을 제공할 것을 의미한다. 따라서 본 연구에 사용된 NR-VMAF는 참조 영상 없이 정확한 사전 품질 예측을 통해 VSR 영상 품질을 효과적으로 보장한다.

VONV의 열 제한 영향 완화 효과를 성능 차이가 큰 두 기기(Galaxy Note 20, Galaxy S23)에서 검증한 결과, 열 제한 발생 시점을 각각 3.16배, 2.9배 지연시키며 열 제한의 영향을 완화함을 확인했다. 또한 NR-VMAF를 기반으로 적응적인 SR을 수행함으로써 전체 추론 시간을 최소 1.4배, 최대 4배까지 단축하였다. 결론적으로 VONV는 NR-VMAF를 활용한 정교한 품질 예측과 동적 VSR을 통해 On-Device VSR에서 품질과 성능 최적화의 균형을 실현하였다.

## V. 결론

On-Device VSR은 DNN 모델의 강한 연산 부하로

인한 발열과 그로 인한 성능 변동성이 중요한 문제다. 이에 대한 해결책인 동적 VSR은 성능 최적화와 영상의 품질의 균형을 안정적으로 유지해야 한다. VONV는 NR-VMAF를 활용한 사전 품질 예측을 기반으로 최적의 SR 방법을 선택하고 On-Device VSR 성과와 영상의 품질을 효과적으로 최적화한다. 본 논문에서 제시한 NR-VMAF 모델은 원본 영상을 참조한 FR-VMAF와 비교하여 높은 정확도와 함께 유의미한 품질 예측 효과를 입증했다. 성능 차이가 명확한 Galaxy Note 20, Galaxy S23에서 실험한 결과 VONV는 열 제한 시점을 각각 3.16배, 2.9배 지연시키고 전체 추론 시간을 최소 1.4배에서 최대 4배까지 단축하였다.

VONV는 기기 상태와 추론 시간 중심의 기존 동적 SR 방식을 벗어나 VMAF 예측을 통한 사전 품질 평가 기반의 동적 SR 방식으로, VMAF 활용의 새로운 접근법을 제시한다. 현재 VONV의 동적 VSR에서는 두 가지 SR 방식만 사용하지만 SR 방법을 세밀화한다면 더 효과적인 최적화가 가능할 것이다. 더불어, 온도, 품질, 시간 등 중점적으로 고려하고자 하는 기준에 따라 적응적인 알고리즘을 추가한다면 다양한 상황에서 최적의 성능을 달성하는 VSR 시스템으로의 발전이 기대 가능하다.

## References

- [1] M. V. Conde, Z. Lei, W. Li, C. Bampis, I. Katsavounidis, and R. Timofte, "AIM 2024 challenge on efficient video super-resolution for AV1 compressed content," *arXiv preprint arXiv:2409.17256*, Mar. 2025. (<https://doi.org/10.48550/arXiv.2409.17256>)
- [2] Netflix, *VMAF - Video Multi-Method Assessment Fusion* (2018), Retrieved Dec. 14, 2024, from <https://github.com/Netflix/vmaf>
- [3] A. De Decker, J. De Cock, P. Lambert, and G. Van Wallendael, "No-reference VMAF: A deep neural network-based approach to blind video quality assessment," in *IEEE Trans. Broadcasting*, vol. 70, no. 3, pp. 844-861, Sep. 2024. (<https://doi.org/10.1109/TBC.2024.3399479>)
- [4] S. Park, Y. Cho, H. Jun, J. Lee, and H. Cha, "OmniLive: Super-resolution enhanced 360° video live streaming for mobile devices," in *Proc. 21st Annual Int. Conf. Mobile Syst., Appl., and Services (MobiSys)*, pp. 261-274, Helsinki, Finland, Jun. 2023. (<https://doi.org/10.1145/3581791.3596851>)
- [5] H. Park and R. Ha, "Optimizing on-device VSR performance via adaptive thermal management," *J. KICS*, vol. 50, no. 1, pp. 1-8, Jan. 2025. (<https://doi.org/10.7840/kics.2025.50.1.1>)
- [6] H. Sheikh and A. Bovik, "Image information and visual quality," in *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, Feb. 2006. (<https://doi.org/10.1109/TIP.2005.859378>)
- [7] S. Li, F. Zhang, L. Ma, and K. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," in *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 935-949, Oct. 2011. (<https://doi.org/10.1109/TMM.2011.2152382>)
- [8] M. Manohara, A. Moorthy, J. De Cock, I. Katsavounidis, and A. Aaron, *Optimized Shot-Based Encodes: Now Streaming!* (2018), Retrieved Dec. 15, 2024, from <https://netflixtechblog.com/optimized-shot-based-encodes-now-streaming-4b9464204830>
- [9] E. Zamfir, Z. Wu, N. Mehta, Y. Zhang, and R. Timofte, "See more details: Efficient image super-resolution by experts mining," in *Proc. 41st ICML*, Vienna, Austria, Jul. 2024. (<https://doi.org/10.48550/arXiv.2402.03412>)
- [10] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, and K. Lee, "NTIRE 2019 challenge on video deblurring and super resolution: Dataset and study," in *Proc. 2019 IEEE/CVF Conf. CVPRW*, Long Beach, CA, USA, Jun. 2019. (<https://doi.org/10.1109/CVPRW.2019.00251>)
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. CVPR*, pp. 770-778, Las Vegas, NV, USA, Jun. 2016. (<https://doi.org/10.48550/arXiv.1512.03385>)
- [12] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th ICML*, pp.

6105-6114, Long Beach, CA, USA, Jun. 2019.  
(<https://doi.org/10.48550/arXiv.1905.11946>)

- [13] J. Ozer, “*Finding the just noticeable difference with Netflix VMAF*,” (2017), Retrieved Dec. 23, 2024, from <https://streaminglearningcenter.com/codecs/finding-the-just-noticeable-difference-with-netflix-vmf.html>
- [14] E. Agustsson and R. Timofte, “NTIRE 2017 challenge on single image super-resolution: Dataset and study,” in *Proc. 2017 IEEE/CVF Conf. CVPRW*, Honolulu, HI, USA, Jul. 2017. (<https://doi.org/10.1109/CVPRW.2017.150>)
- [15] Y. Kim and D. Son, “Noise conditional flow model for learning the super-resolution,” in *Proc. 2021 IEEE/CVF Conf. CVPRW*, Nashville, TN, USA, Jun. 2021. (<https://doi.org/10.48550/arXiv.2106.04428>)

박 혜 진 (Hye-jin Park)



2025년 2월 : 홍익대학교 컴퓨터공학과 졸업  
2025년 3월~현재 : 연세대학교 컴퓨터공학과 석사과정  
<관심분야> 기계학습, 온디바이스 컴퓨팅, 뉴럴 네트워크  
[ORCID:0009-0006-3000-8638]

하 란 (Rhan Ha)



1987년 : 서울대학교 컴퓨터공학 학사  
1989년 : 서울대학교 컴퓨터공학 석사  
1989년~1990년 : KT 전임연구원  
1995년 : University of Illinois at Urbana-Champaign 컴퓨터공학 박사  
1995년~현재 : 홍익대학교 컴퓨터공학과 교수  
<관심분야> 머신 러닝, 온디바이스 컴퓨팅, 사물인터넷, 인터미턴트 컴퓨팅, 임베디드 시스템, 모바일 컴퓨팅, 실시간 시스템, SW보안  
[ORCID:0000-0001-9861-362X]