AI 기술 설명서

1. 프로젝트 개요

[프로젝트 명칭: 실시간 웃음판별 딥러닝 모델]

본 프로젝트는 사용자의 미세한 표정 변화를 실시간으로 감지하여 '웃음' 여부를 공정하게 판정하는 **고성능 경량 AI 심판**을 개발하는 것을 목표로 합니다.

---

2. 핵심 기술 구성

1. MobileNetV3-Small (CNN) : 이미지 탐색
2. Bi-LSTM : 시퀀스 특징 추출
3. Dual-Head 구조 : 웃음 여부(이진 분류)뿐만 아니라 현재 사용자의 감정 상태(다중 분류)를 동시에 분석
4. Masked Multi-task Learning : Dual-Head 구조를 한번에 학습하기 위한 방식

---

3. 모델 아키텍처 상세

3.1 데이터 입력 및 전처리

입력 텐서 구조: `(B, T, 3, 224, 224)`

- B: Batch Size
- T: Sequence Length (시퀀스 길이)
- (3, 224, 224): RGB 채널 및 이미지 해상도

Reshape: CNN 백본은 시간 정보를 직접 처리하지 못하므로, 입력 텐서를 `(B * T, 3, 224, 224)`로 재구성하여 모든 프레임을 개별 이미지로 처리

3.2 MobileNetV3-Small Backbone

Depthwise Separable Convolution을 사용하여 연산량을 획기적으로 절감

연산량 비교

- 일반 Convolution 연산량:
  $Cost_{sep} = D_K^2 \cdot M \cdot N$
- Depthwise Separable 연산량:
  $Cost_{sep} = (D_K^2 \cdot M) + (M \cdot N)$

Feature Projection: 576차원의 벡터를 256차원으로 압축하여 LSTM이 처리

3.3 시간적 맥락 분석 (Bi-LSTM Layer)

1프레임 화면보다, 시간적 변화를 포착할 때 정확도가 상승할 것으로 기대

Bidirectional(양방향) 구조를 사용하여 특정 시점의 표정을 앞뒤 프레임의 맥락과 함께 분석

3.4 다중 작업 결정 (Dual-Head Structure)

LSTM의 마지막 은닉 상태를 두 개의 독립적인 Fully Connected Layer로 분기합니다.

- **Head 1 (Smile Detection)**: 이진 분류를 통해 웃음 발생 확률(0~1)을 출력
- **Head 2 (Emotion Classification)**: 다중 분류를 통해 현재 사용자의 감정 상태를 예측

3.5 Masked Multi-task Learning

손실 함수

- Smile Detection: Sigmoid + Binary Cross Entropy
- Emotion Classification: CrossEntropyLoss

Masking & Total Loss 계산:

- 최종 손실값은 다음과 같이 각 Head의 가중치(w)를 곱하여 합산:

$$Loss_{total} = (w_{smile} \cdot Loss_{smile} \cdot Mask_{smile}) + (w_{emotion} \cdot Loss_{emotion} \cdot Mask_{emotion})$$

정답이 없는 데이터(라벨 -1)의 오차를 0으로 만드는 **Masking 연산**이 포함

```
# target이 -1인 곳을 찾아서 0과 1로 구성된 마스크 생성
mask = (target != -1).float()

# 원본 오차 계산
raw_loss = criterion(pred, target)

# 마스크 곱하기 -> -1이었던 부분의 loss는 0이 됨
masked_loss = raw_loss * mask

# 최종 평균 오차 (0이 된 부분은 제외하고 평균 계산)
final_loss = masked_loss.sum() / (mask.sum() + 1e-6)
```

---

4. 데이터셋

4.1 GENKI-4K

기본 정보

- 샘플 수: 4,000 이미지
- 클래스: 2가지 (Smiling, Non-smiling)

4.2 SMILES Dataset

기본 정보

- 샘플 수: 13,165 이미지
- 클래스: 2가지 (Smiling: 3,690 / Non-smiling: 9,475)
- 해상도: 64×64 그레이스케일

4.3 FER2013

기본 정보

- 샘플 수: 35,887 이미지
- 클래스: 7가지 (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral)
- 해상도: 48×48 그레이스케일

4.4 CelebA (Smiling Attribute)

기본 정보

- 샘플 수: 202,599 이미지
- 인물 수: 10,177명
- 속성: 40가지 binary attributes (Smiling 포함)
- Smiling 라벨: 각 이미지에 웃음 여부 표시

---

5. 한국인 특화 Fine-tuning

기존 데이터셋의 서양인 편향성을 극복하기 위해 직접 수집한 한국인 특화 데이터셋으로 Fine-tuning을 진행할 예정. 실제 팀원들이 화상통화에 참여하여 웃거나 안 웃는 녹화본을 촬영한 후, 프레임을 나눠 학습 데이터로 사용할 계획.

---

6. 기대 효과

본 모델은 실시간 웹캠 환경에서 경량화된 구조로 빠른 추론 속도를 제공하면서도, 시간적 맥락을 고려한 정확한 웃음 판별을 수행합니다. 특히 한국인 데이터로 Fine-tuning을 진행함으로써, 기존 모델 대비 한국인 표정 인식 정확도를 크게 향상시킬 수 있을 것으로 기대됩니다.