2 공모 제안서 양식-실현 가능 서비스 부문

※ 글씨 크기 10pt, 서체 맑은 고딕으로 통일하여 작성해 주세요.
※ 실제 데이터가 개방되어 있지 않은 경우는 반드시 데이터별 첨부된 '테크니컬 리포트'를 참고하여 아이디어를 제안해 주세요.
※ 이미지, 동영상 등 자료 첨부 시, 본인이 저작권을 가지고 있는 자료를 사용하거나, 본인의 저작권이 없는 경우 반드시 저작권자 출처를 명시해 주세요.(URL포함)
※ 제안서 양식의 내용은 어디까지나 참고 자료입니다. 제안 시 자유롭게 아이디어를 제안해 주세요.

1. 인공지능 학습용 데이터 활용 아이디어 제목

콘텐츠 맞춤 지능형 수어 통역 영상 제작 솔루션

2. 인공지능 학습용 데이터 활용 아이디어 내용

● **서비스 제안 동기**

- 최근 TV 뉴스 프로그램 이외에, 대국민 담화나 주요 정책, 재난 상황 발표 현장 등의 방역 대책을 위한 **공공 분야에서도 알 권리와 언어권 보장을 위해 수화 통역이 제공.**
- 다양한 분야에서 수화 통역 서비스를 제공하려고 노력 중이지만, 전국 기준 40만 명의 농인들을 위한 **통역사는 600명 수준**으로 매우 열악한 상황.
- 또한, 수화 통역 영상 제작을 위해선 별도의 스튜디오에서 촬영과 편집이 요구되어 **인력 문제와 함께 시간과 비용의 문제가 발생.** 이를 대체하기 위해 3D 애니메이션 제작[1], 영상 편집 등의 방법이 제안되고 있지만, 수어 특성상 **표정과 자세의 구현 정확도가 떨어져** 실제 활용도가 낮음.

● **서비스 실현 가능성**

- 2016년 **심층 합성곱 생성적 적대 신경망 (DCGAN: Deep Convolutional Generative Adversarial Networks)**가 제안된 이후, GAN 관련 연구와 기술은 매우 빠르게 성장 중
- 최근에는 GAN 기술이 정교해지며, 패션 디자인 및 가상 모델 생성, 흑백 이미지 복원, 변조 영상 판별 기법 등 산업 비즈니스 모델과 융합 ⇨ **산업 분야에서의 활용 가능성이 높게 평가됨**
- 현재 GAN은 충분한 데이터와 학습조건만 형성된다면 고해상도 영상을 정교하게 구현 가능한 수준까지 도달하였으며, 단순히 얼굴 외형 이외에도 다양한 표정이나 모습까지 표현됨

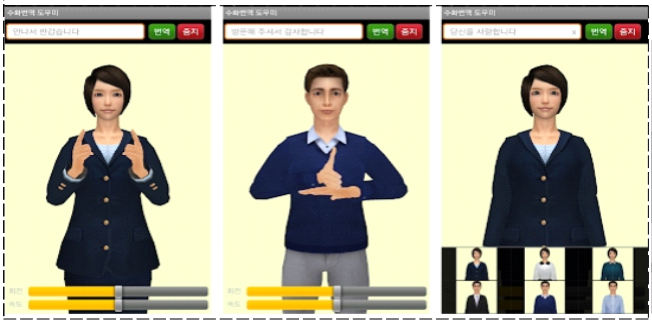

그림 1. 수화 번역 및 통역 서비스[1]

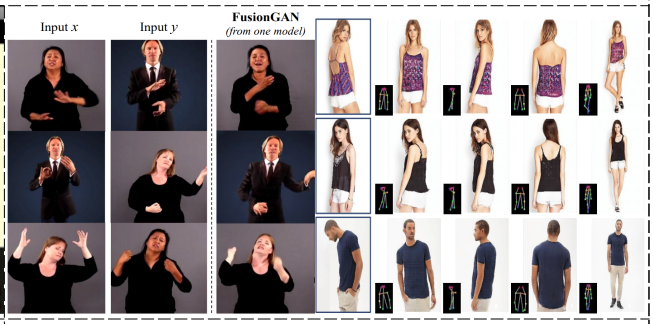

그림 2. GAN을 통해 생성된 사람의 표정 및 외형[2-3]

● **서비스 개요**

- **GAN 기반 영상 재생성 및 SWAP 기술을 이용한 콘텐츠 맞춤 수화 통역 영상 제공**
- **통역 서비스가 필요한 영상 콘텐츠의 통역을 제공하는 비즈니스 모델의 솔루션**
- **학습 모델 구축 이후 한국어 이외 다양한 문화권 언어 적용 가능성 확보**

● **서비스 제공 절차**

- 영상 대본을 수화 통역을 위한 1차 전처리 가공 [음절 분리, 조사, 관사 등 불필요 텍스트 제거]
- 가공된 텍스트를 입력하여 단일 단위의 수어 영상 데이터 정렬 [GAN 입력 영상 1차 전처리]
- 기존 영상 콘텐츠에 맞게 싱크 조절 및 가공 [GAN 입력 영상 2차 전처리]
- 가공된 입력 영상과 재생성을 위한 기존 데이터를 이용하여 GAN 학습 모델 구축 [GAN 학습]
- **후처리 작업으로 생성된 영상의 배경이나 크로마키 등을 설정 및 최종 서비스 제공**

● **서비스 목표**

- 수화는 단순히 청각 장애인을 위한 서비스가 아닌 알권리를 보장받기 위한 하나의 언어체계.
- 제안하는 서비스 실현을 통해 **농인들의 알권리 보장과 더불어, 다양한 시청각 콘텐츠 활용하고 소비의 폭을 넓힐 수 있도록 활용**되는 것을 최종 목표

그림 3. 서비스 제공 절차

3. 아이디어를 실현하기 위해 필요한 인공지능 학습용 데이터

● 본 서비스는 표정과 행동 데이터를 기반으로 수화 영상을 재생성하는 GAN 신경망 알고리즘이 핵심. 사람의 행동을 좌표화한 데이터와 얼굴 랜드 마크 좌표 데이터 그리고 수어 영상 데이터가 요구됨 ⇨ **수어 데이터는 수집이 어렵고 전문성이 높은 데이터라 확보가 어려우나 AI 허브 데이터 구축 사업으로 인해 아이디어 실현 가능성 확보**

● **활용된 인공지능 학습용 데이터**

- 수어에 따른 표정 및 자세 데이터 확보를 위한 **수어 영상 데이터**
- GAN 기반 자세 재생성 모델 구축을 위한 **사람 동작 영상 데이터** 일부
- 테스트 및 구현을 위해 실내 공간에서 전신 영상으로 수집된 **YouTube Pose 데이터**

4. 인공지능 학습용 데이터 학습방법

● **관절 키포인트 탐지/예측 알고리즘 학습**

- 관절 키포인트 예측 모델은 현재 Top-Down (사람 탐지 후 관절 예측)과 Down-Top (관절 위치 예측 후 사람 탐지) 방식으로 분류되며 응용 환경에 따라 프레임워크가 다르게 설계됨.
- 제안하는 서비스는 1인 통역사를 기준으로 예측/재생성하는 모델 ⇨ Down Top 방식 모델 선정.
- 학습 데이터 추출을 위해 CNN Down-Top 방식의 **2D Pose Estimation**[4] (CFA, PRM)를 활용하여 수어 영상 데이터에서 분류된 클래스 **(left/right hand 21개, pose 25개)**의 포인트를 학습.
- 학습 성능 고도화를 위해 모든 수어 데이터는 촬영된 모든 각도의 영상 데이터를 입력으로 학습하며 K-Fold 교차검증 방식으로 수어 데이터에 대한 성능 평가 결과 검증.
- 모델 구축 후, **사람 동작 영상과 YouTube Pose 영상 데이터를 기반으로 모델 성능을 추가 평가**.

● **얼굴 랜드마크 (키포인트) 탐지 알고리즘 학습**

- 관절 키포인트 탐지 알고리즘 학습과 마찬가지로 **CNN 방식으로 적용되는 Facial Landmark Detection 알고리즘**[5]을 활용하여 수어 영상에서 분류된 **클래스 (face 68개)의 포인트 학습.**
- 이후 모델 성능 평가 방식은 관절 키포인트 탐지 알고리즘 학습과 동일하게 적용.

● **GAN 신경망 알고리즘 학습**

- GAN은 **생성자**와 **판별자**의 경쟁 학습으로 최적화되는 학습 메커니즘을 가짐.⇨ 입력되는 수어 영상과 학습으로 완성할 모델에서 추출된 영상이 함께 입력되어야 함.
  [학습된 모델 (관절/얼굴)을 적용하여 재생성을 원하는 데이터에서 라벨 생성
- GAN 융합 모델을 기반으로 한 포즈 재생성 **Liquid Warping GAN**[6] 알고리즘으로 영상 재생성
- 세부적으로 표현되지 못한 얼굴은 **FACE/POSE SAWP 알고리즘 FSGAN**[7]으로 재생성 및 보정
- **공개되는 수어 데이터 영상의 특성 및 라벨 분포도에 따라 융합 GAN 신경망 응용**

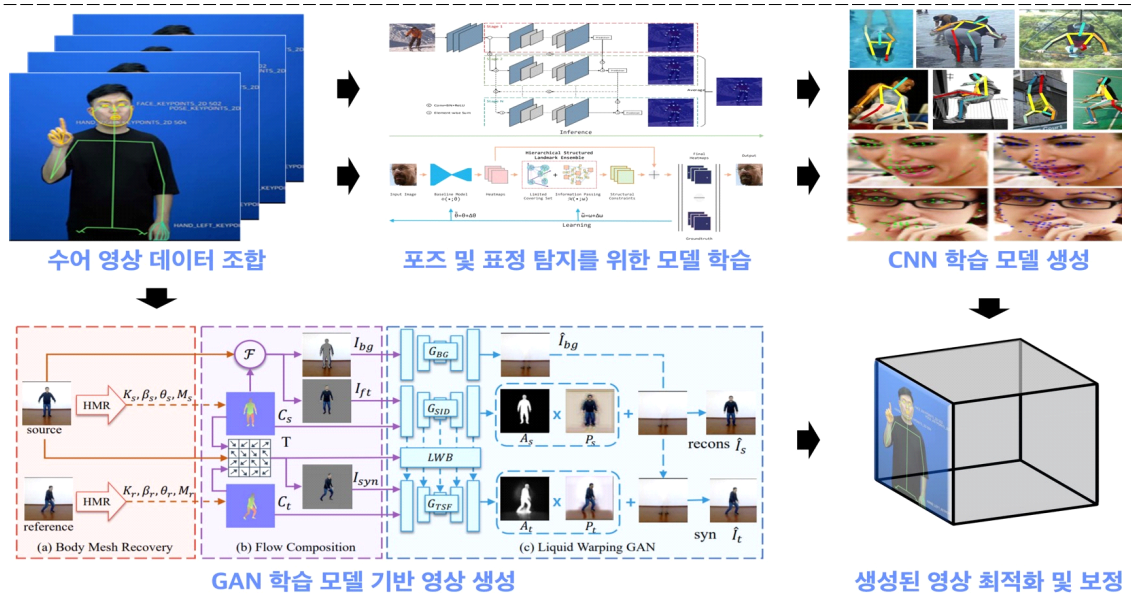

그림 4. 전체 학습 네트워크 과정 예시[4-6]

5. 참고자료

1. **삼성전자 수화도움 번역기**
   https://play.google.com/store/apps/details?id=com.sec.android.app.ksldic&hl=ko&gl=US
2. **Generating a Fusion Image: One's Identity and Another's Shape**
   https://arxiv.org/abs/1804.07455
3. **Generating Person Images with Appearance-aware Pose Stylizer**
   https://arxiv.org/abs/2007.09077
4. **Cascade Feature Aggregation for Human Pose Estimation**
   https://arxiv.org/abs/1902.07837
5. **Learning Robust Facial Landmark Detection via Hierarchical Structured Ensemble**
   https://par.nsf.gov/servlets/purl/10161302
6. **Liquid Warping GAN: A Unified Framework for Human Motion Imitation, Appearance Transfer and Novel View Synthesis**
   https://arxiv.org/abs/1909.12224
7. **FSGAN: Subject Agnostic Face Swapping and Reenactment**
   https://arxiv.org/abs/1908.05932