IPTC-Amazon collaboration

Meeting

(January 23rd, 2023)

POLITÉCNICA

iptc

www.iptc.upm.es

IPTC-Amazon: Index

Index:

- Sign language motion generation from high level sign characteristics
- Speaker diarization with multimodal inputs
- Pose and spatial movement as input for dynamic content search & generation
- Entangling AI-audio synthesis models and multimodal representations
- Zero-shot sonorizing of video sequences

Sign language motion generation and analysis

iptc

POLITÉCNICA

Sign language motion generation and analysis

What have we done?

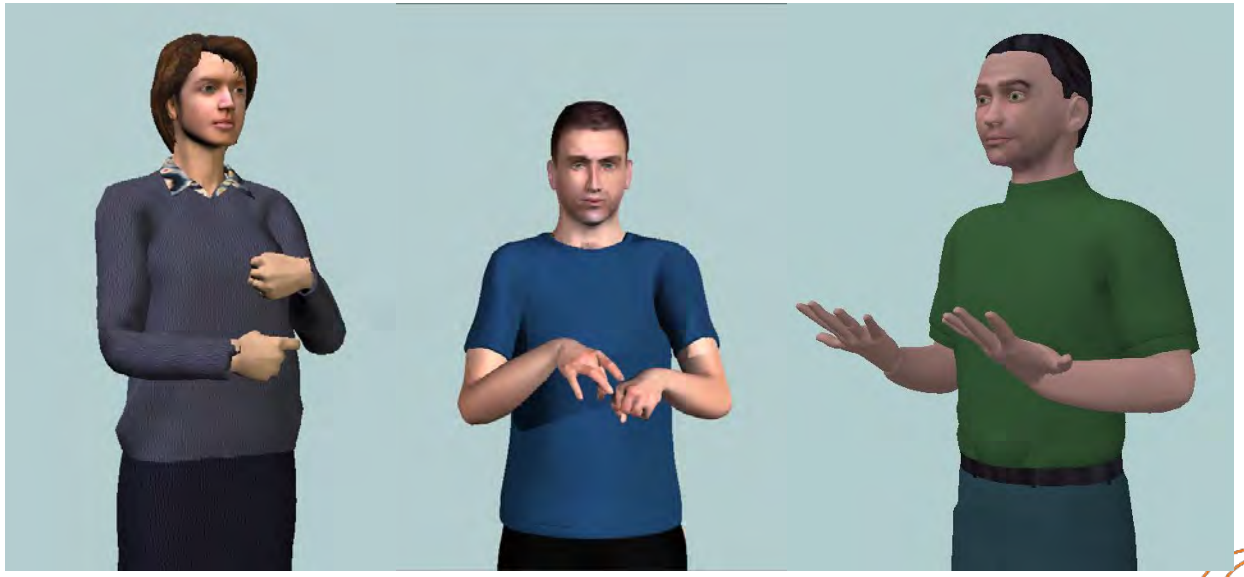
- Generating the Dataset
- Retrieving the entries for the Transformer
- LSTM Project Action Detection
- Technologies Used

iptc

Sign language motion generation and analysis

Generating the Dataset

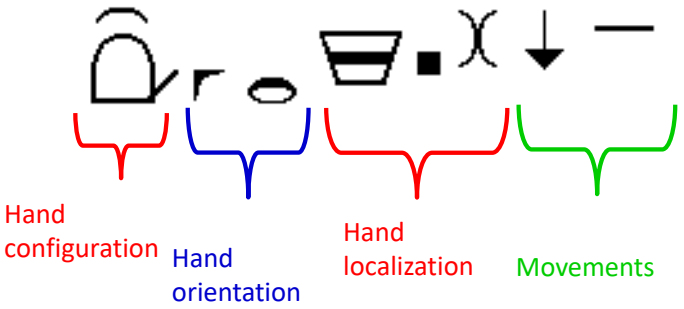
- 750 videos for each avatar and perspective
- 6750 videos in total
- 10-20 frames for each sign

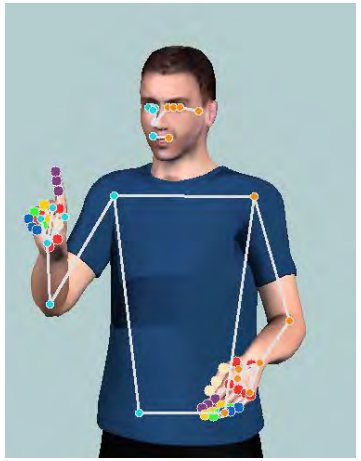


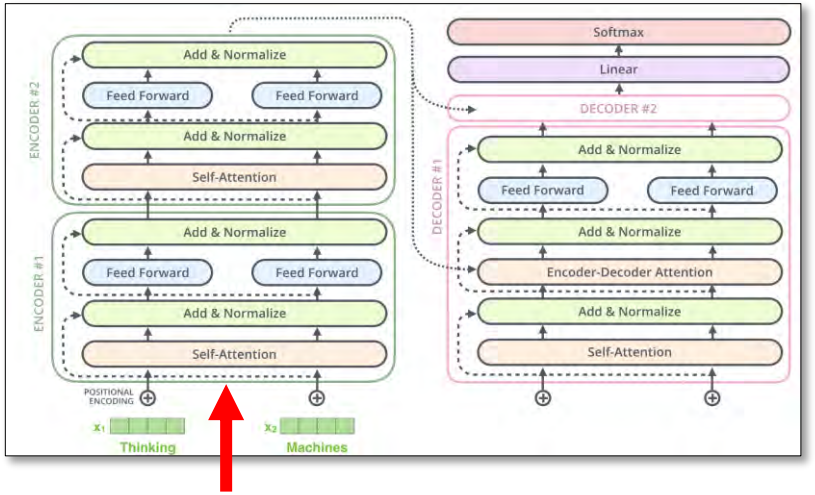
Sign language motion generation and analysis

Next steps

- Review dataset
- Extract x, y, z coordinates
- Extract hand configurations







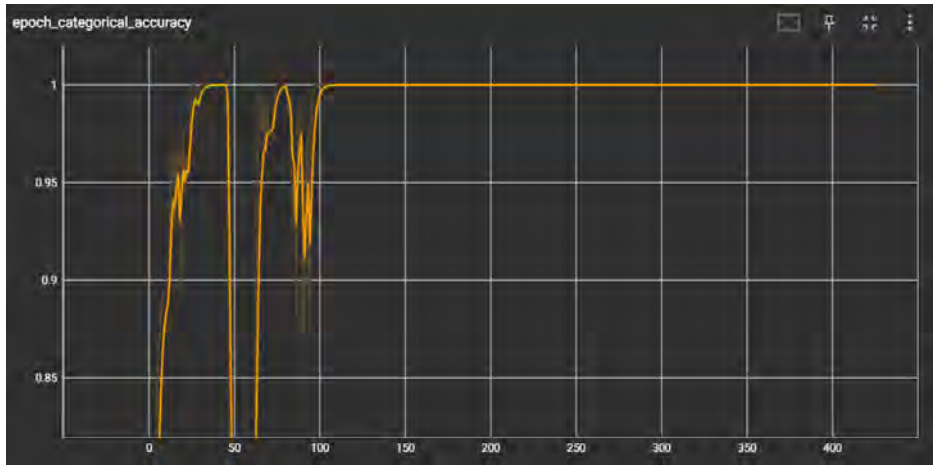
Hand configuration

```
<sigml>
        <hns_sign gloss="$PROD">
                <hamnosys_nonmanual>
                        <hnm_mouthpicture picture="'a"/>
                </hamnosys_nonmanual>
                <hamnosys_manual>
                        <hamfist/>
                        <hambetween/>
                        <hamfist/>
                        <hamthumbacrossmod/>
                        <hamextfingeru/>
                        <hampalmd/>
                        <hamshoulders/>
                        <hamlrat/>
                </hamnosys_manual>
        </hns_sign>
</sigml>
```

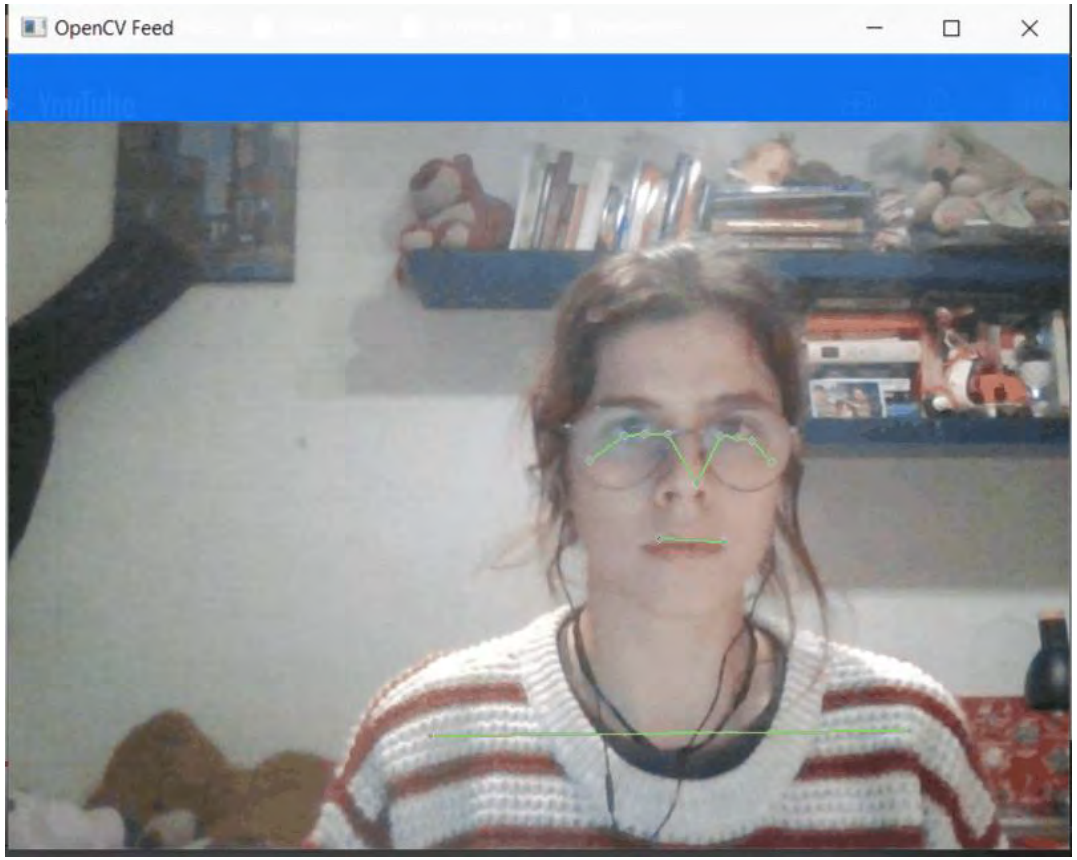
Sign language motion generation and analysis

LSTM PROJECT: ACTION DETECTION

- Action Detection of 3 words
- Optimization function: ADAM
- Loss function: categorical_crossentropy
- Tensorboard: Supervise the model



epoch_categorial_accuracy



https://www.youtube.com/watch?v=doDUihpj6ro

Sign language motion generation and analysis

TECHNOLOGIES USED

- MediaPipe: MediaPipe Holistic to optimize hands and pose components.
- OpenCV: Open-Source plataforma for image processing.
- Tensorflow: Data preprocessing, creation and training of the model. Tensorboard

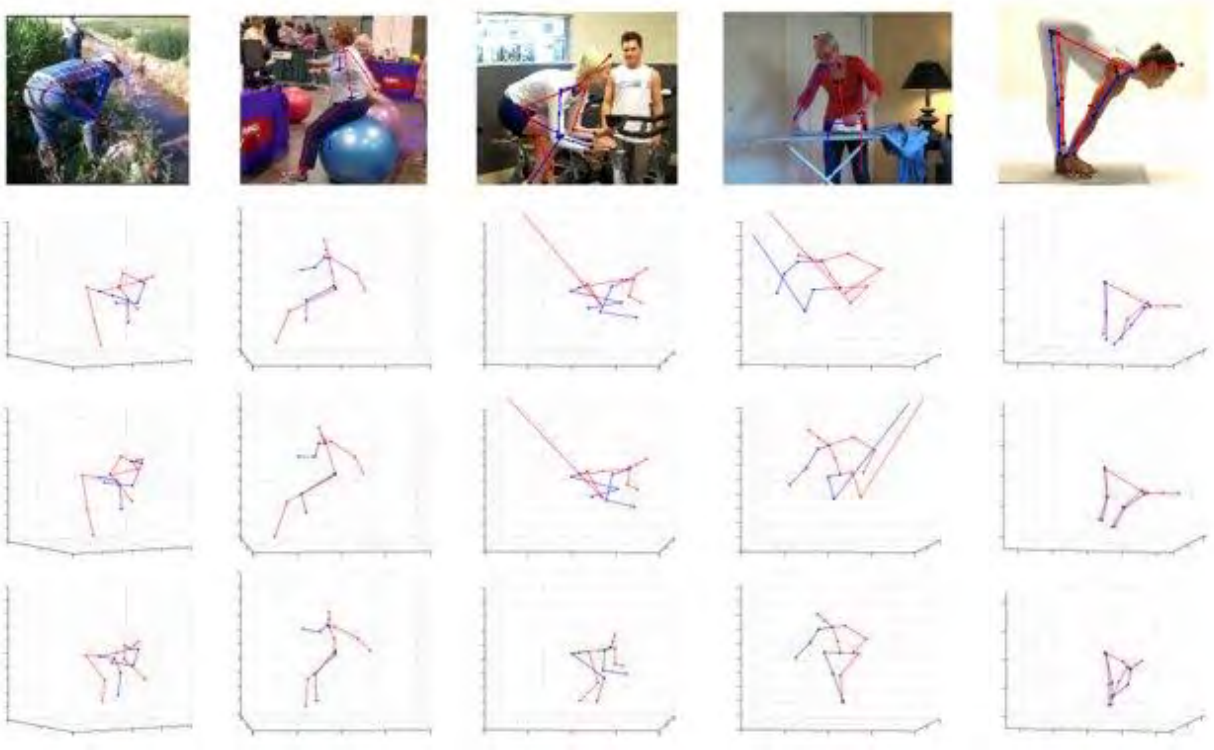
Pose and spatial movement as input for dynamic content search & generation

iptc

POLITÉCNICA

Pose and spatial movement as input for dynamic content search & generation

Main purpose? To explore the potential of posture correctness analysis and multimodal feedback delivery for different applications (ergonomics, yoga, others).



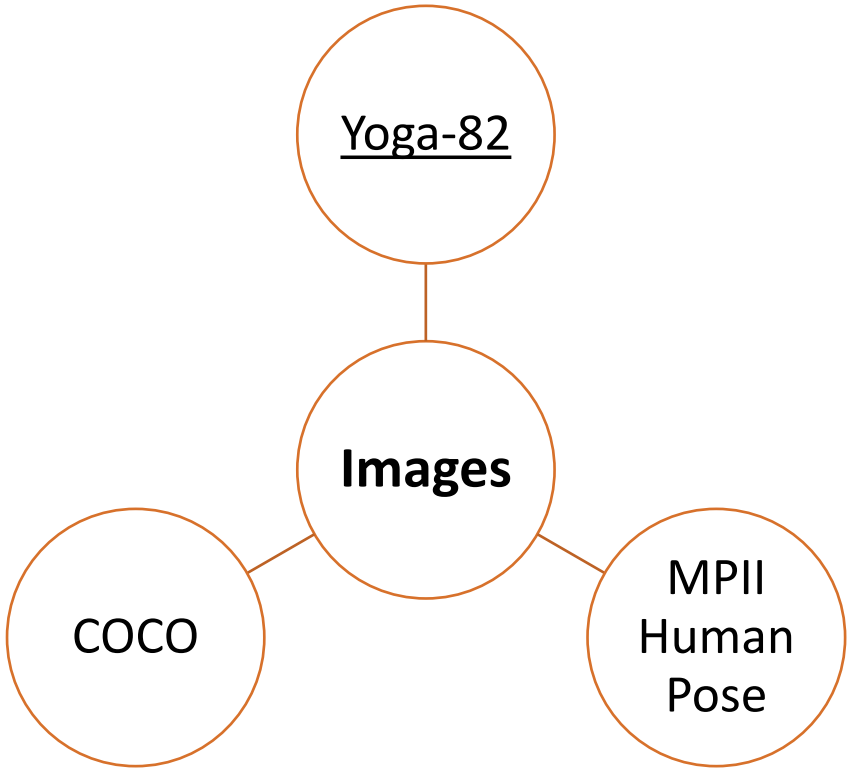
Tasks

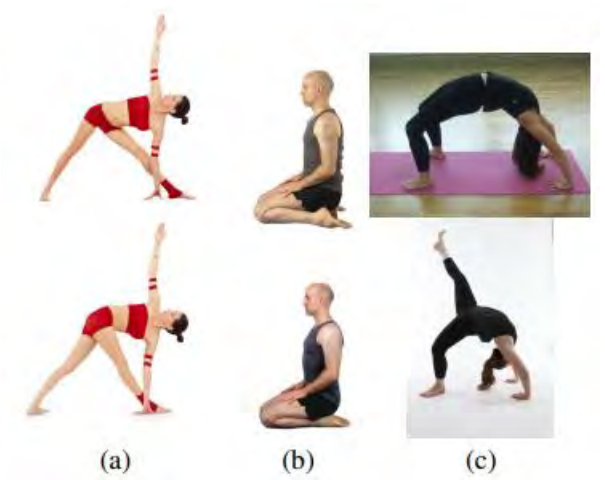
1) Task scope definition
2) Dataset search and evaluation.
3) State of the art on building postural models and postural analysis.
4) Setting up an environment for posture extraction from images/video.
5) Model concept proposal, based on distances and normalizations. Built from reference datasets and literature. Limited scope.
6) Multimodal feedback by using virtual assets (e.g. avatar)
7) Incremental prototype set up.

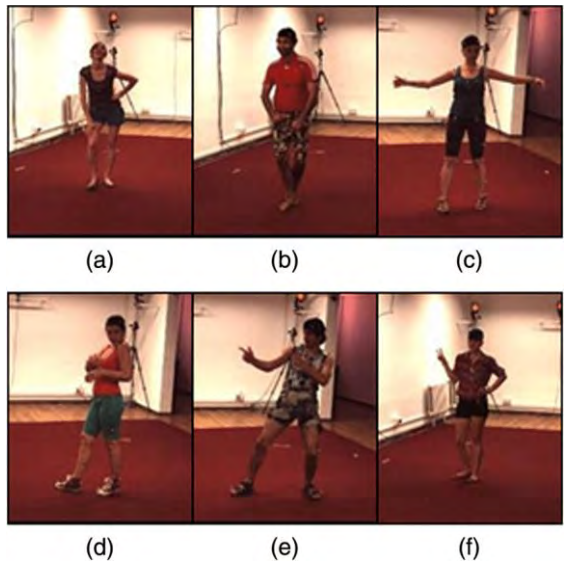
iptc

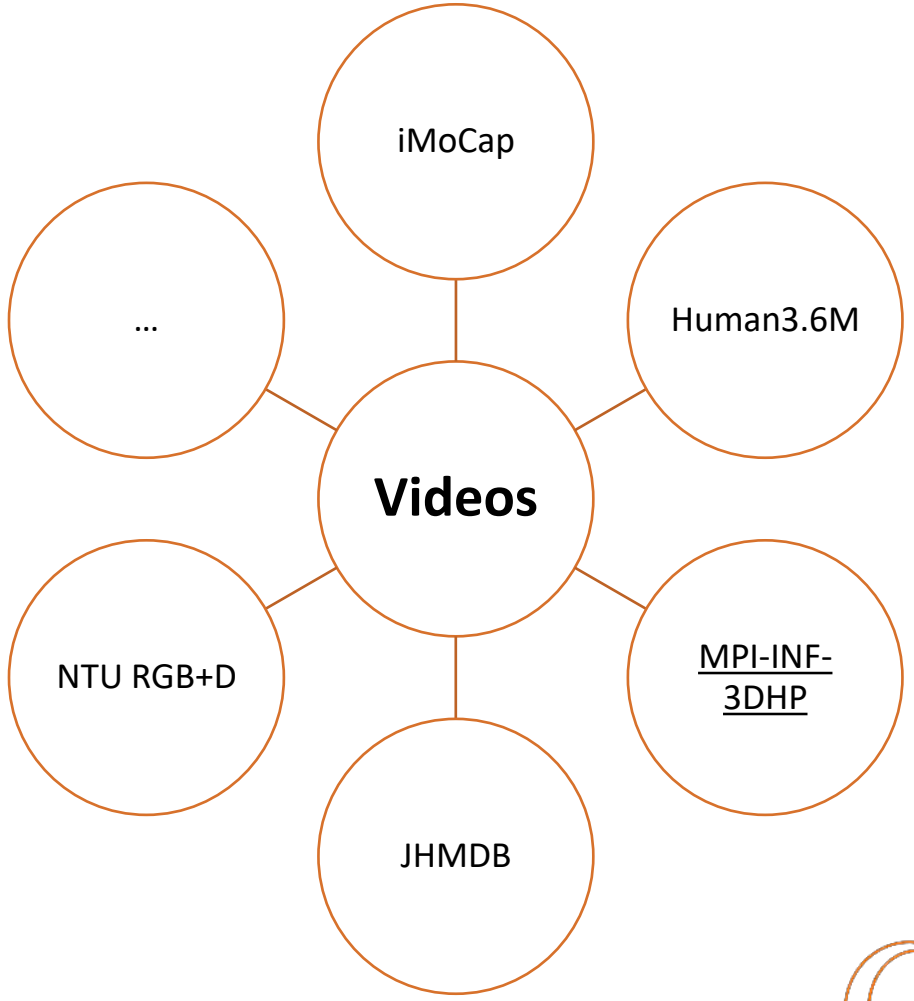
Pose and spatial movement as input for dynamic content search & generation

Analysis of available datasets









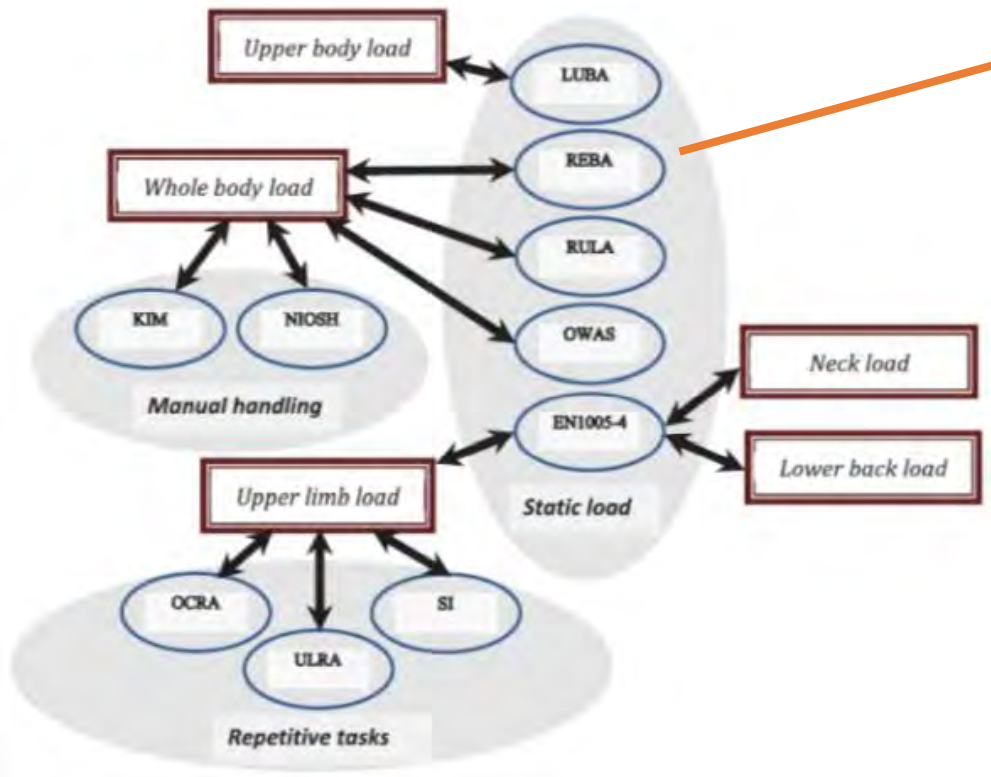
Pose and spatial movement as input for dynamic content search & generation

Postural analysis

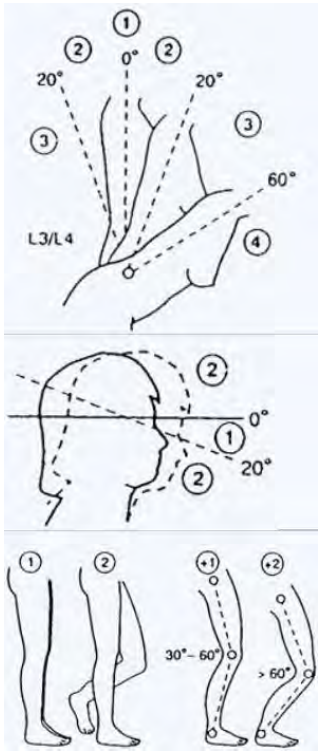
Possible application: How do we compare the input pose with the baseline pose?

- By comparing angles between body axes of the human body and extremities

Methods of ergonomic evaluation



REBA method

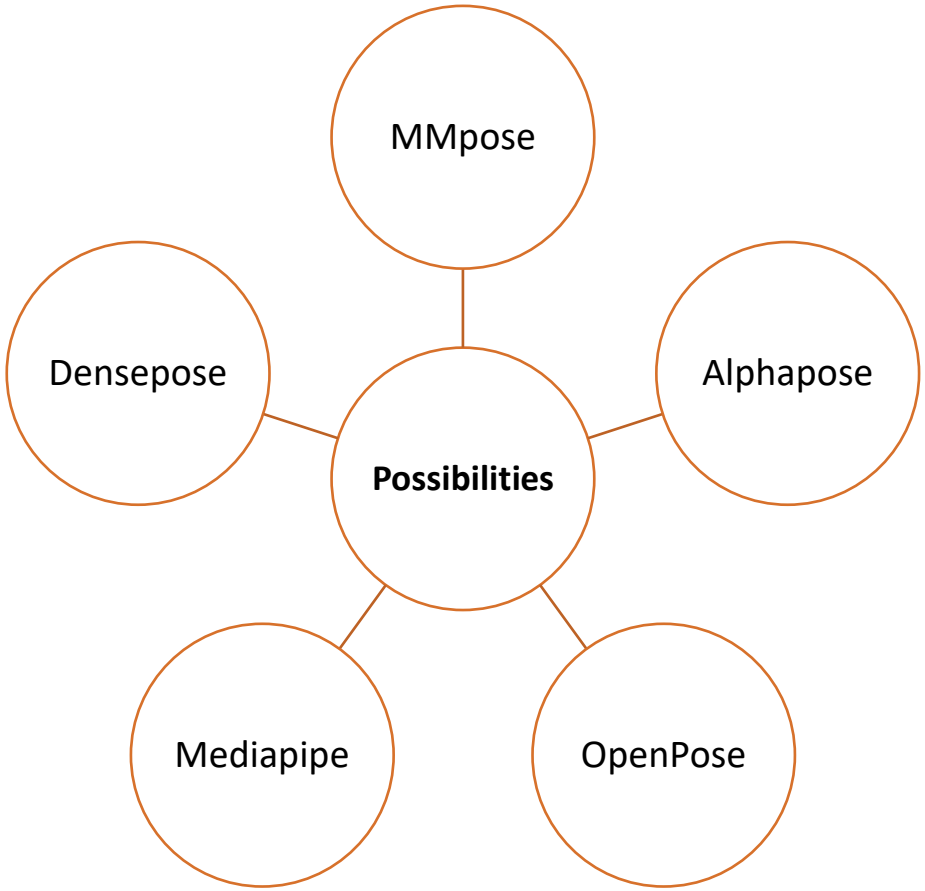


Scoring based on posture and angle

Pose and spatial movement as input for dynamic content search & generation

Next Steps

Checking the capabilities and limitations of each tool



Preliminary analysis:

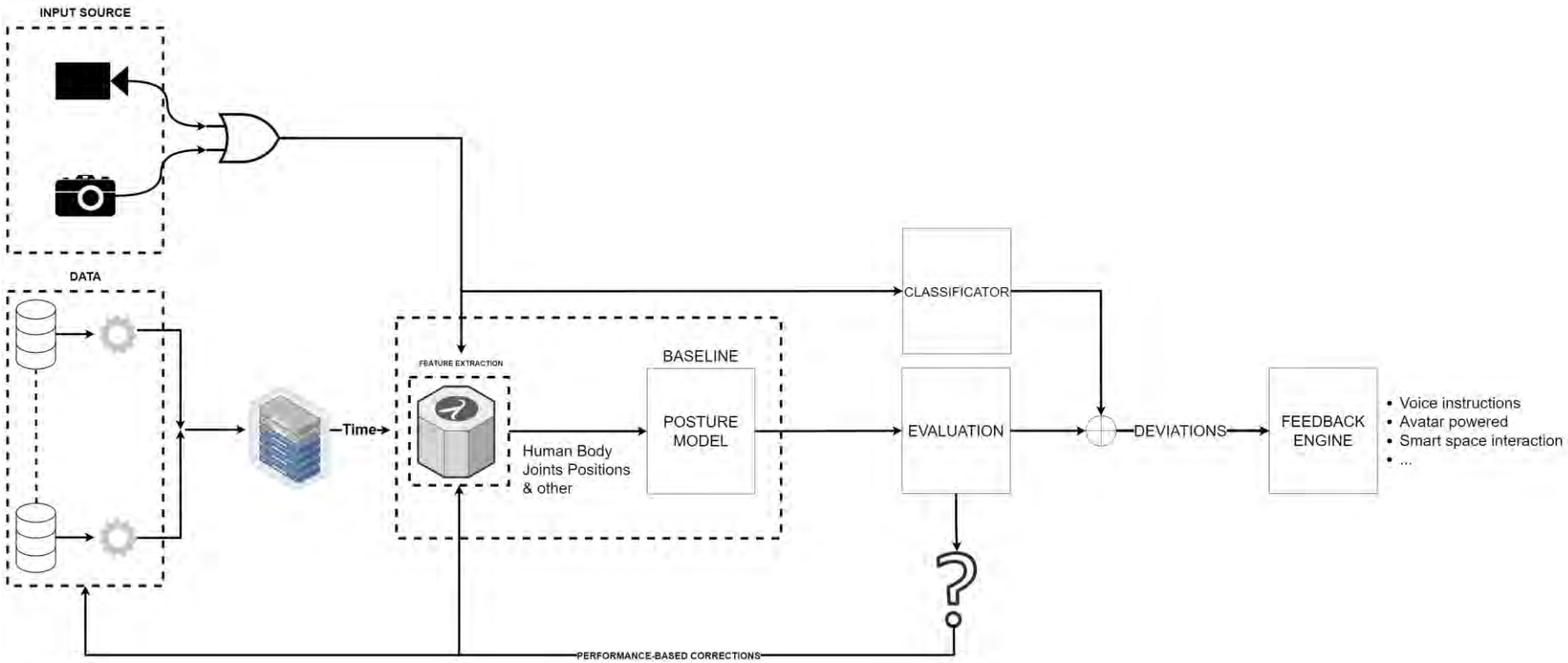
- Comparing between MMpose, OpenPose and AlphaPose, better AlphaPose
- DensePose, no skeleton but mesh points, more difficult association to a physical meaning.
- Mediapipe, a very high resolution in face recognition. Multiplatform!

iptc

Pose and spatial movement as input for dynamic content search & generation

Next Steps

- Set the testing environment
- Searching for the appropriate network architecture
- Transfer learning and use of the dataset selected



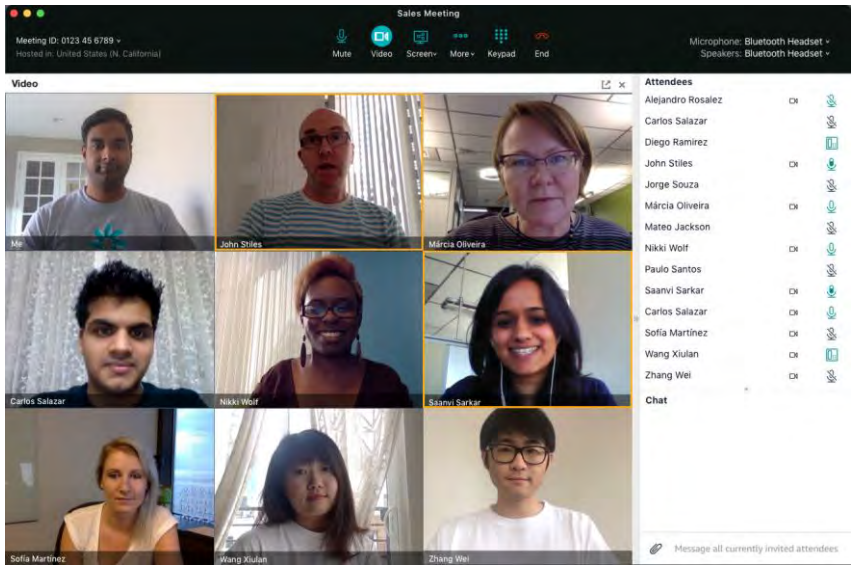
Speaker diarization with multimodal inputs

iptc

POLITÉCNICA

Speaker diarization with multimodal inputs

INTRODUCTION



Video call meetings

Speaker diarization with multimodal inputs

DATASET - AVSpeech

More than 200k videos of duration around 5-10 seconds

Speaker diarization with multimodal inputs

DATASET PREPROCESSING

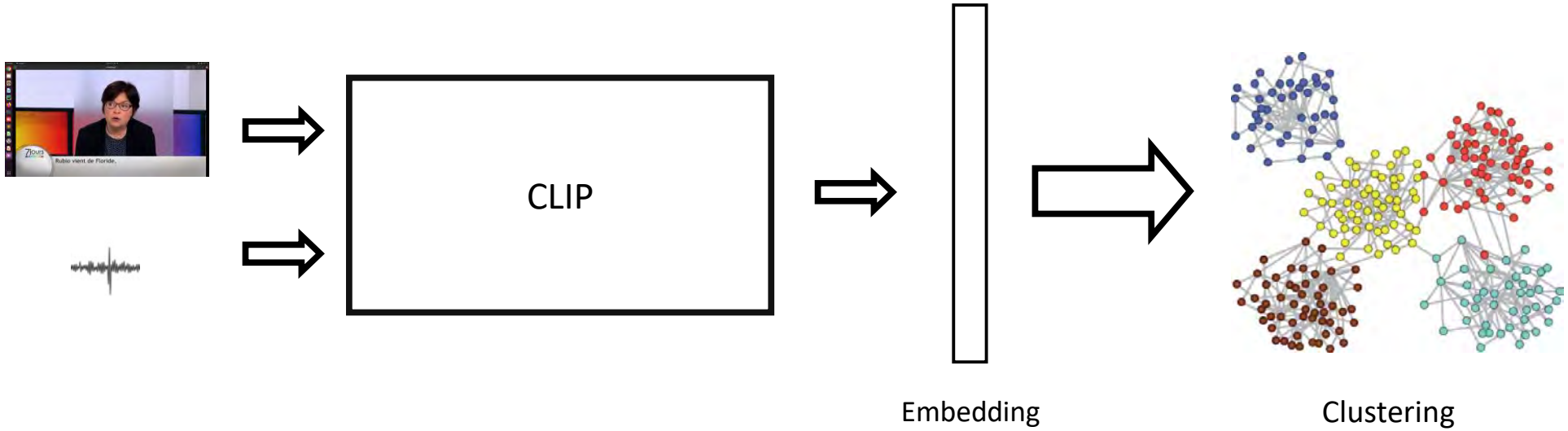
Video

Sub clips

Audio and video

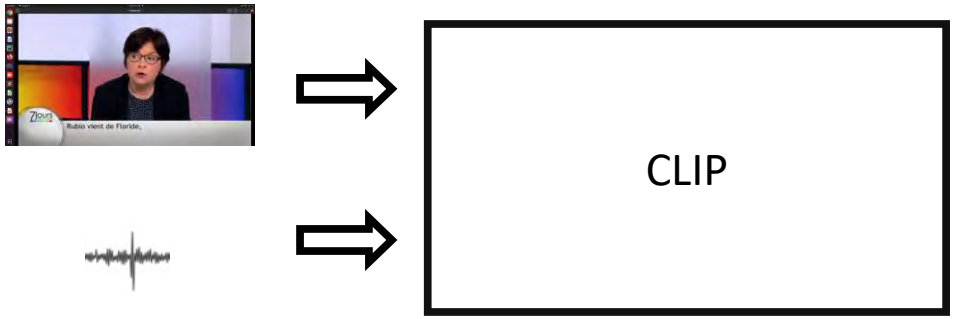
Speaker diarization with multimodal inputs

ARCHITECTURE



Speaker diarization with multimodal inputs

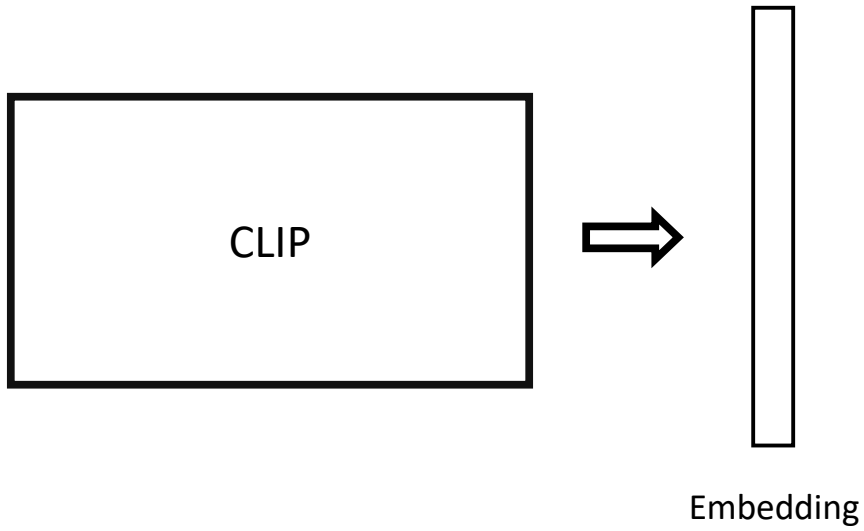
STEP 1: TRAINING



Train the neural network to be able to match a face with a voice

Speaker diarization with multimodal inputs

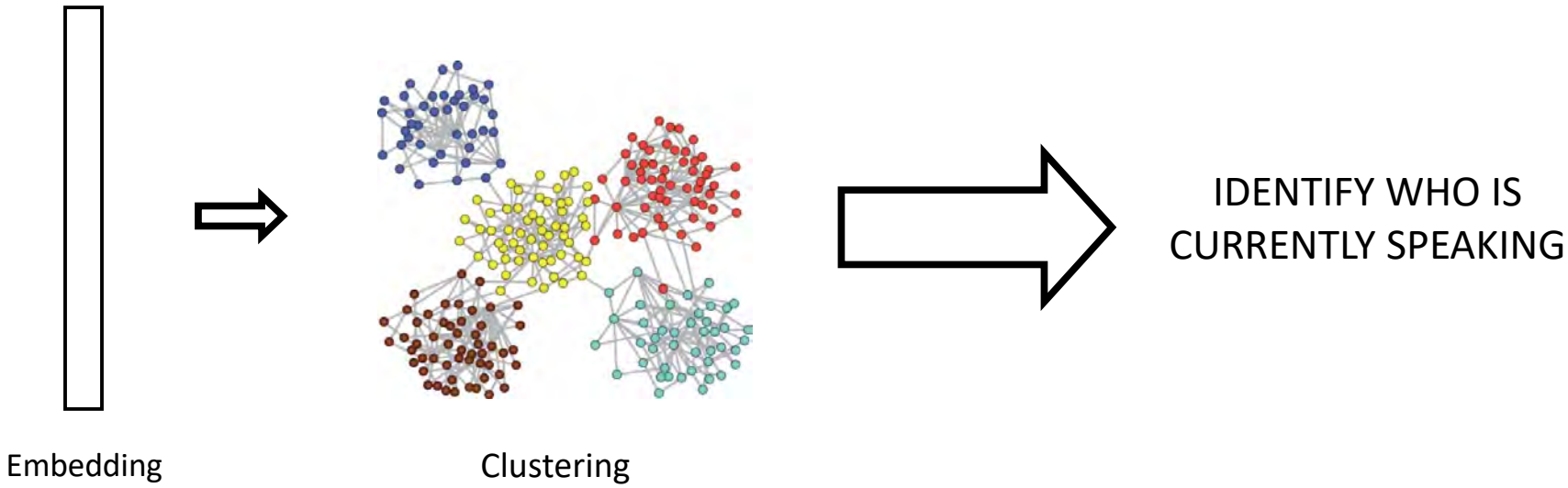
STEP 2: EMBEDDING GENERATION



Generate the embeddings from the input data

Speaker diarization with multimodal inputs

STEP 3: CLUSTERING



Speaker diarization with multimodal inputs

EVALUATION

Audio Generation using Deep Learning

Using AI to generate content

Introduction

- Platforms that generate images:

OpenAI

Dream Studio – Stable Diffusion

A comic book cover of a doctor with huge eyes

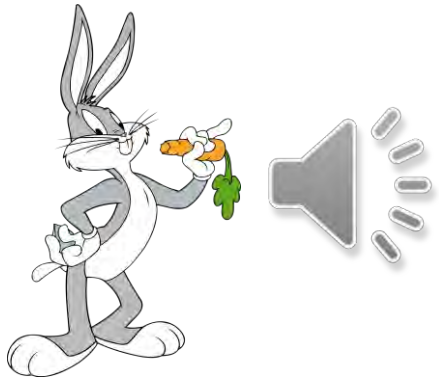
A Shiba Inu dog wearing a beret and black turtleneck

Images generated with Dall·E 2

- Platforms that generate audios:

FakeYou

OpenAI Jukebox



Bugs Bunny generate audio from FakeYou

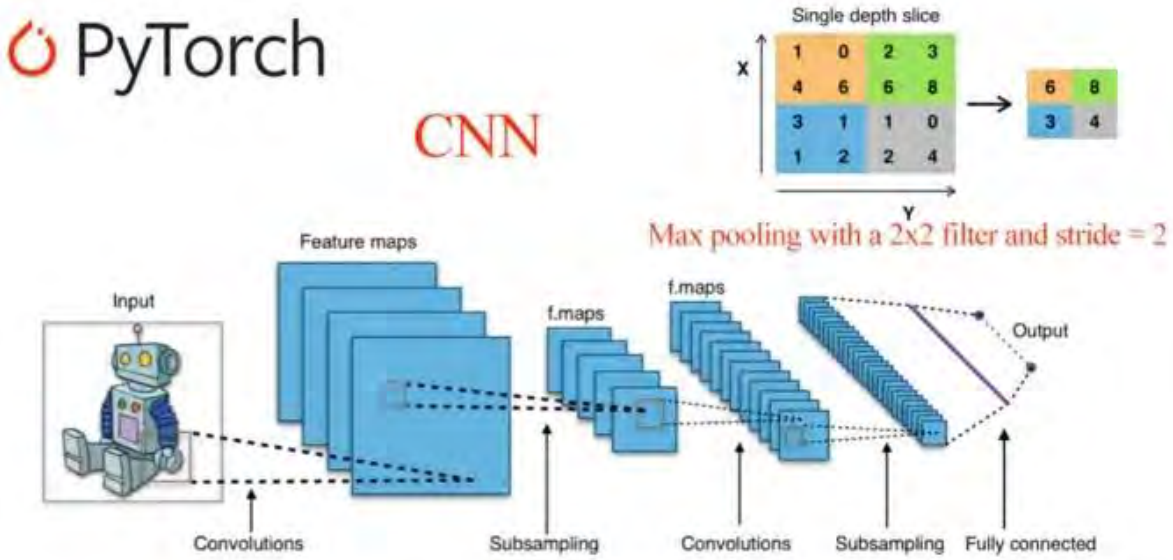
Music made from scratch with OpenAI Jukebox

iptc

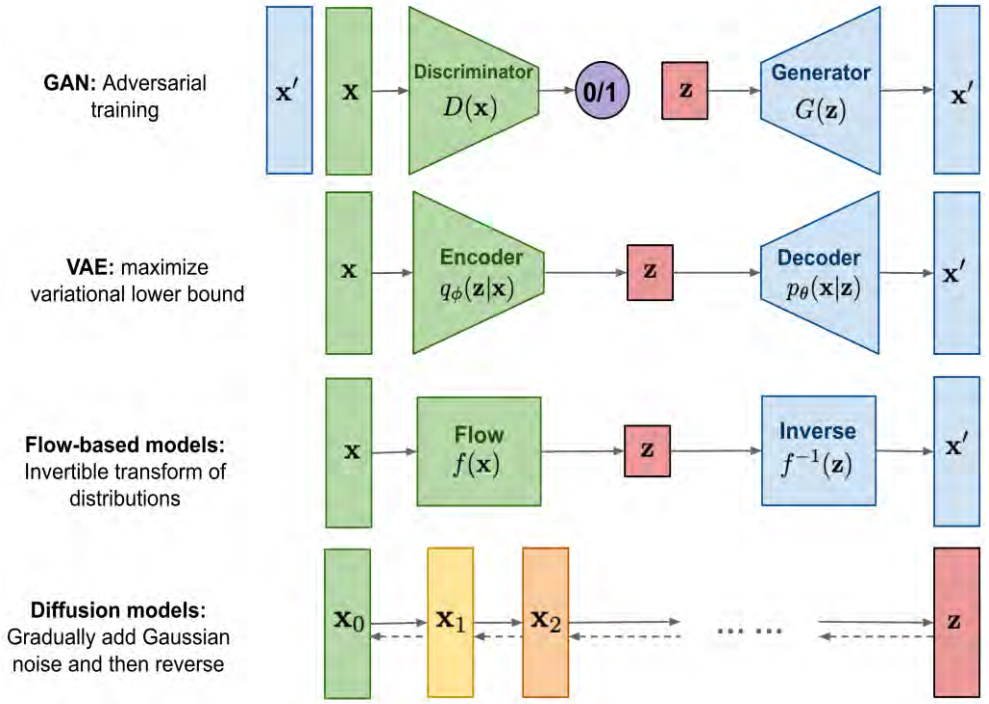
DL Training Seminars

Task 1

CNNs & PyTorch

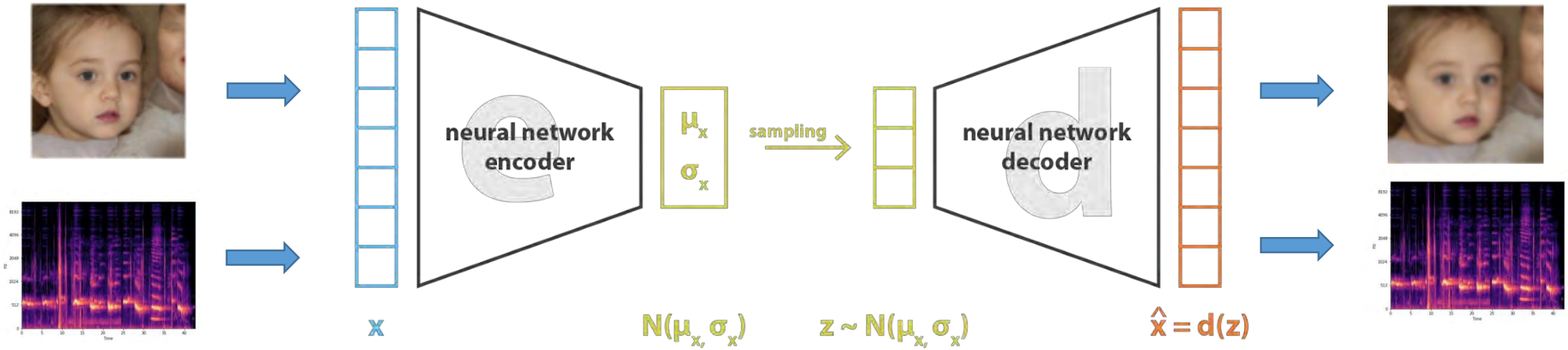


Generative Models



Variational Autoencoder (VAE)

Task 3

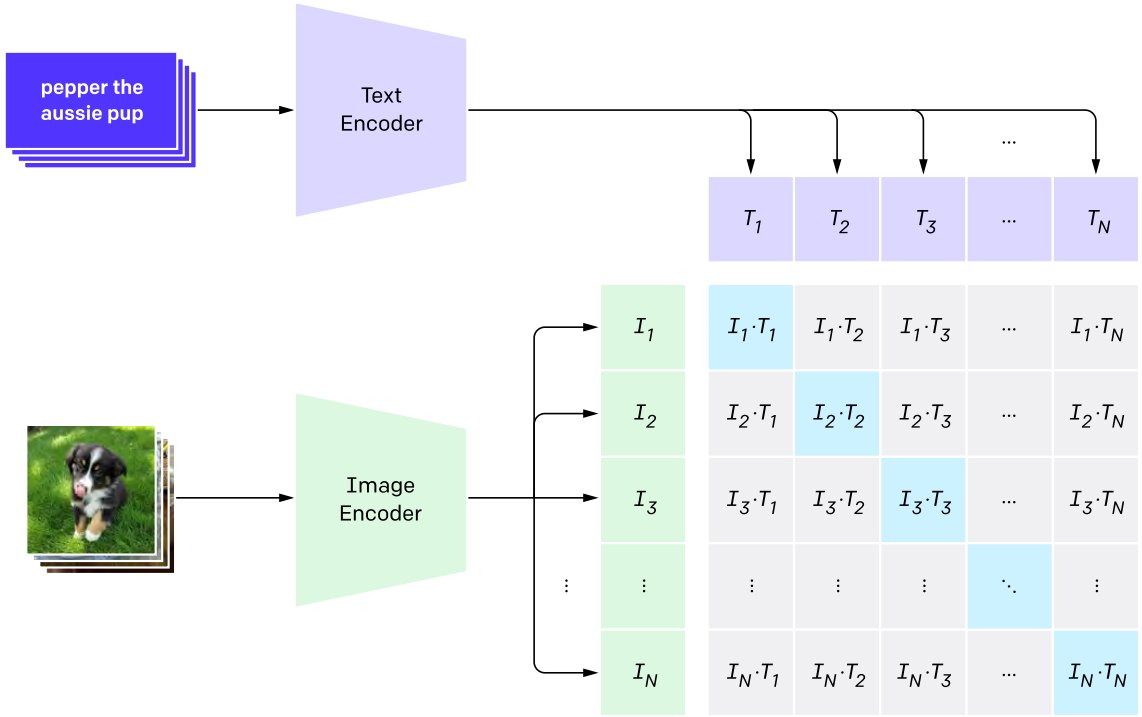


$$\text{loss} = \| x - \hat{x} \|^2 + KL[\, N(\mu_x, \sigma_x), N(0, I) \,] = \| x - d(z) \|^2 + KL[\, N(\mu_x, \sigma_x), N(0, I) \,]$$

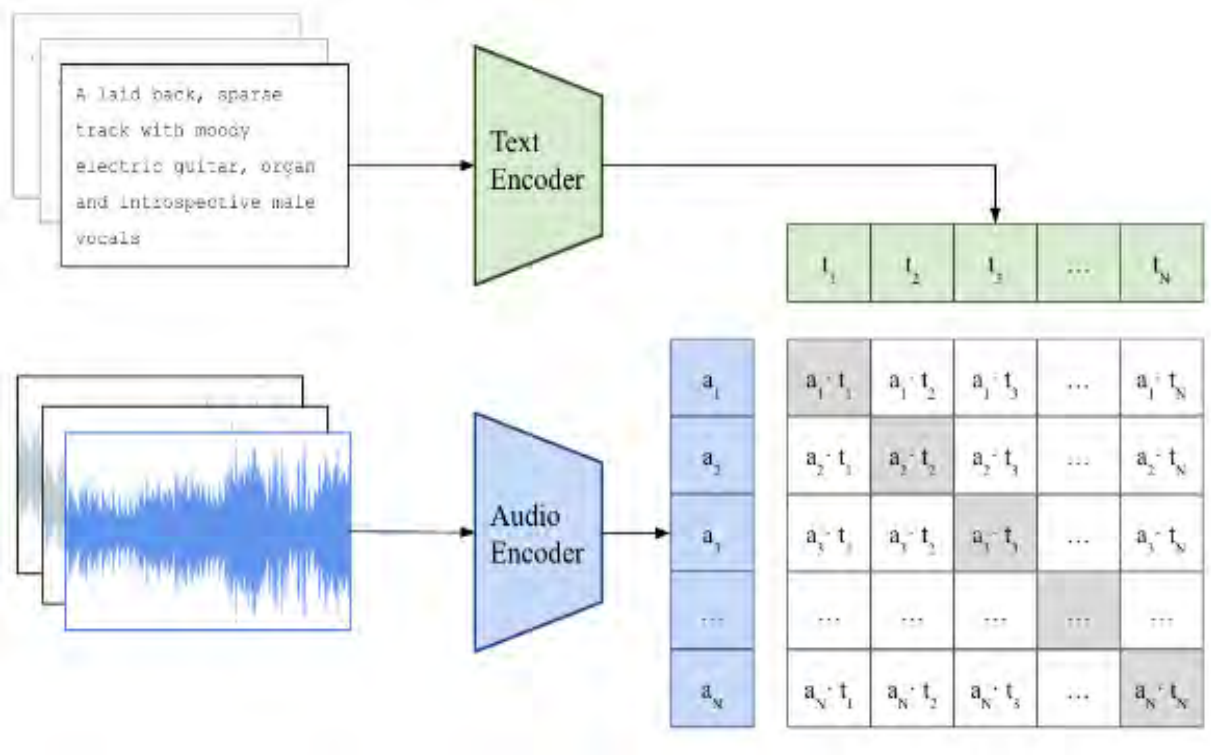
Why CLIP and CLAP?

Task 2

CLIP

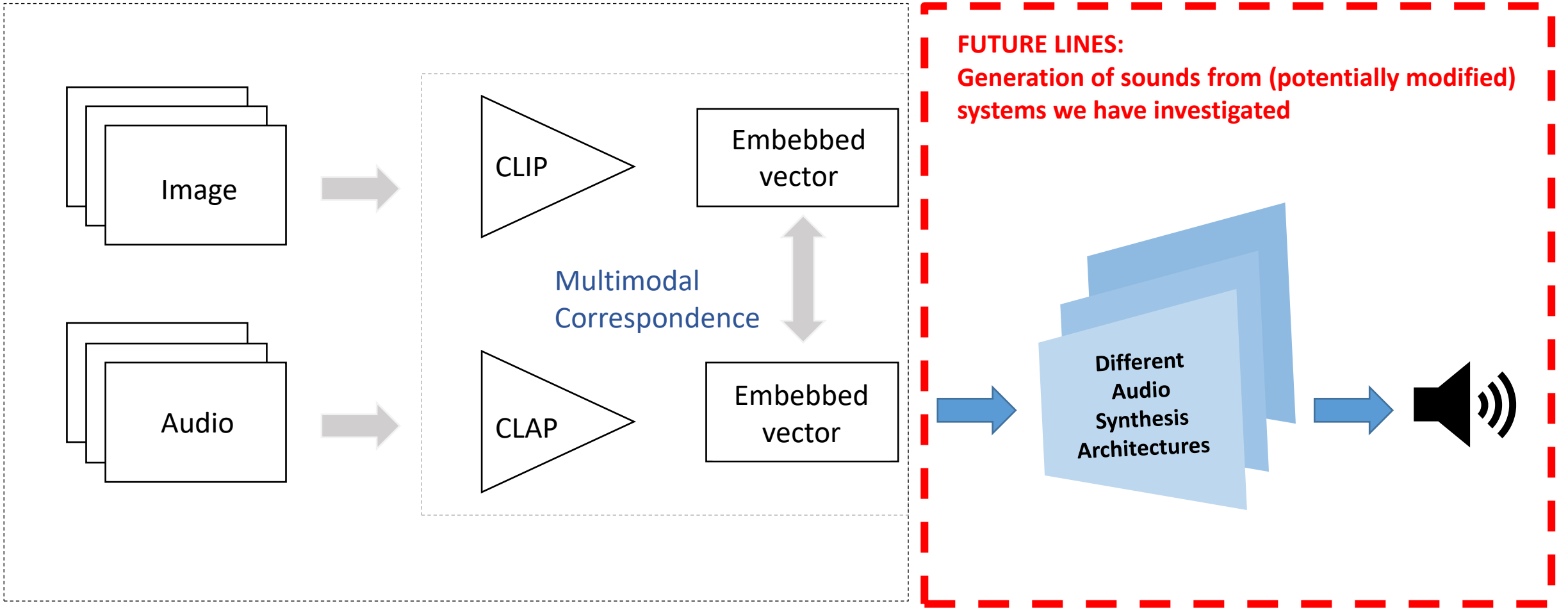


CLAP

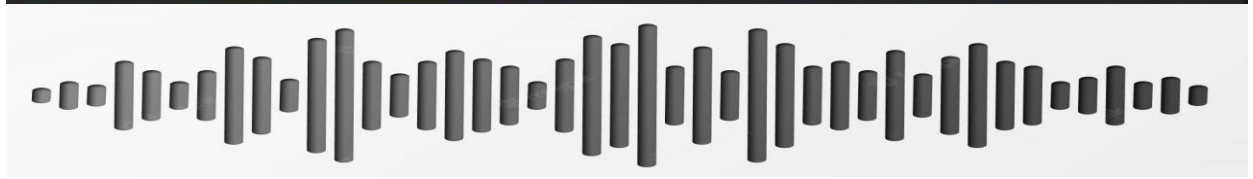


Future Lines in our Projects

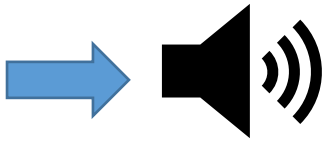
Next Task



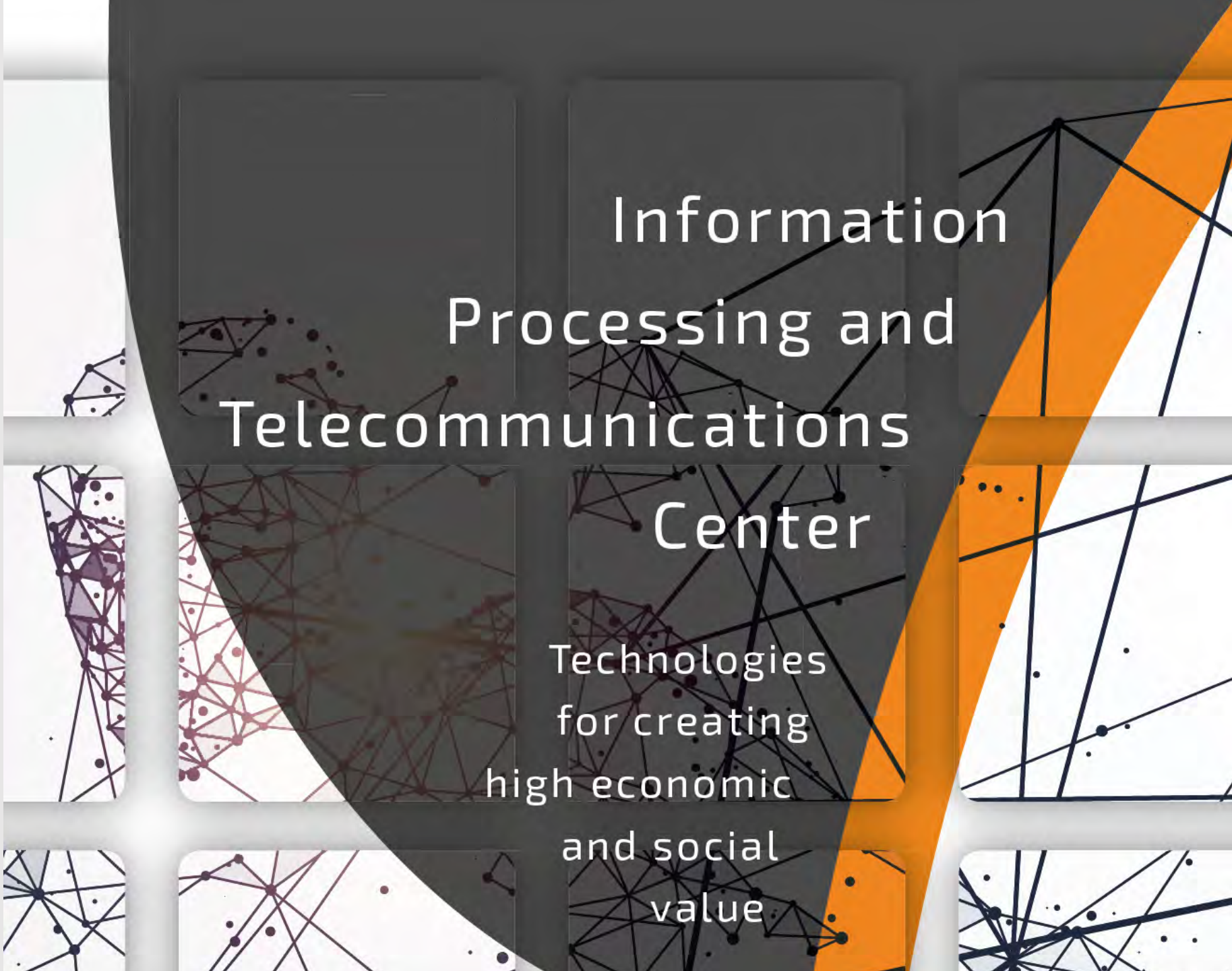
Images, short-videos, video games,… sonorization



Different Audio Synthesis Architectures



Many thanks



POLITÉCNICA

iptc

www.iptc.upm.es