



Εὐρηκα

# 聲不由己

## - 聲音風格轉換

劉至軒

王政祺

陳維超教授

陳怡君老師

彭天建老師

π

# Deepfake

Εύρηκα

π

source:  
Buzzfeed

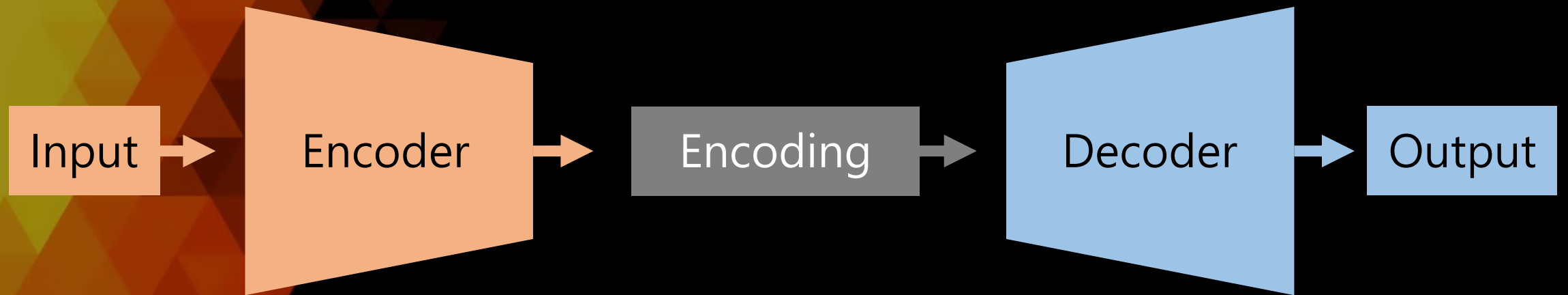
Εύρηκα

# 研究動機 – 新加坡海外旅行



Εύρηκα

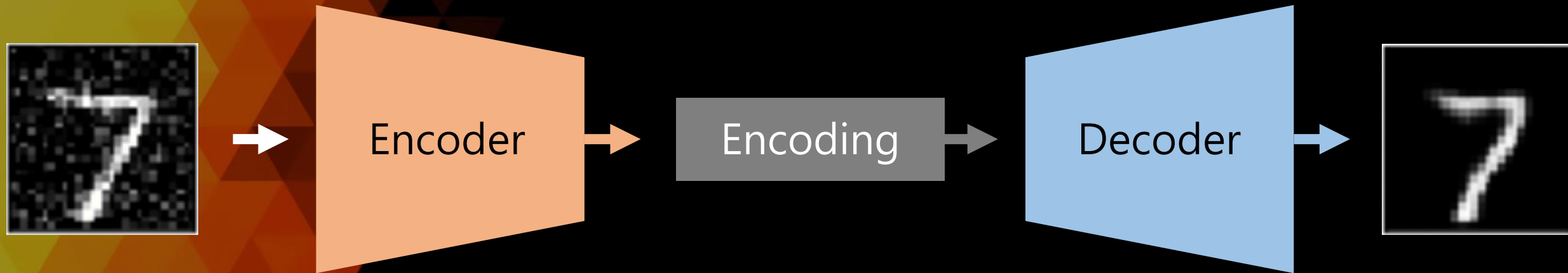
# 理論知識 – Autoencoder



π

Εύρηκα

# 理論知識 – Autoencoder

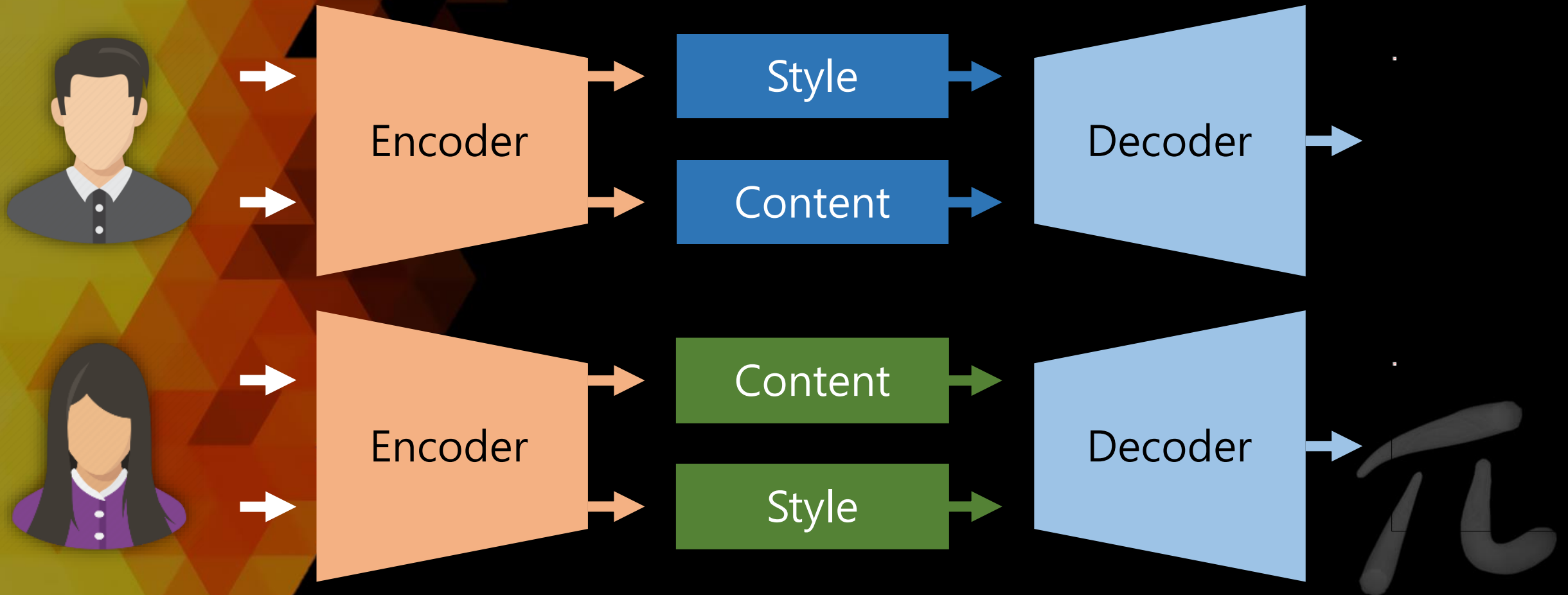


π



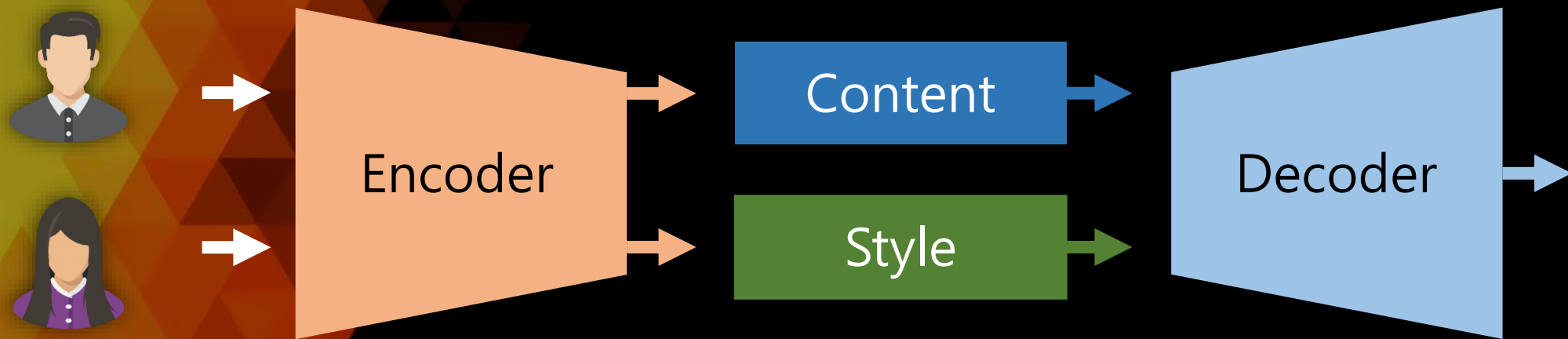
# 理論知識 – AutoVC

Εύρηκα



# 理論知識 – AutoVC

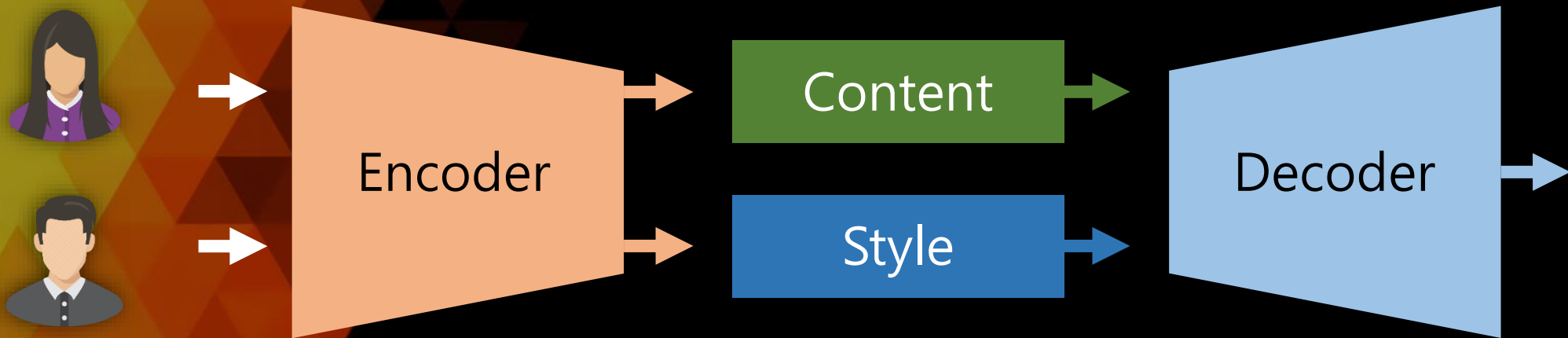
Εύρηκα



π

# 理論知識 – AutoVC

Εύρηκα

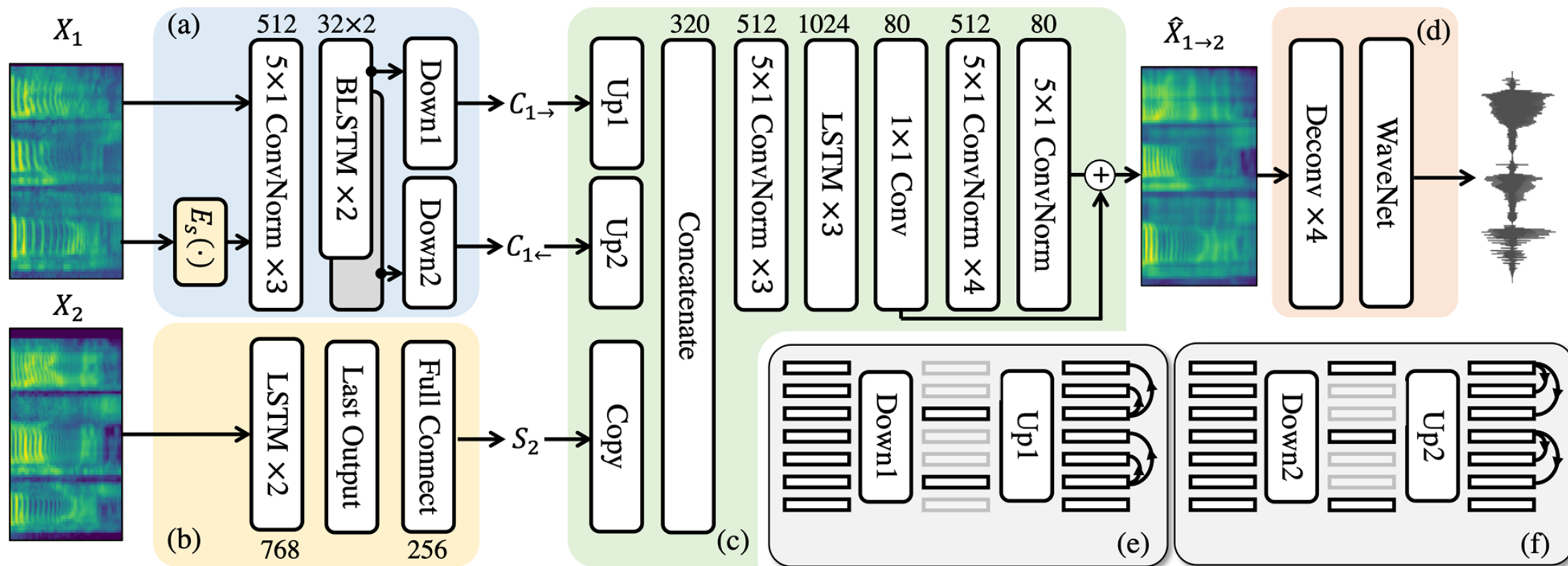


π



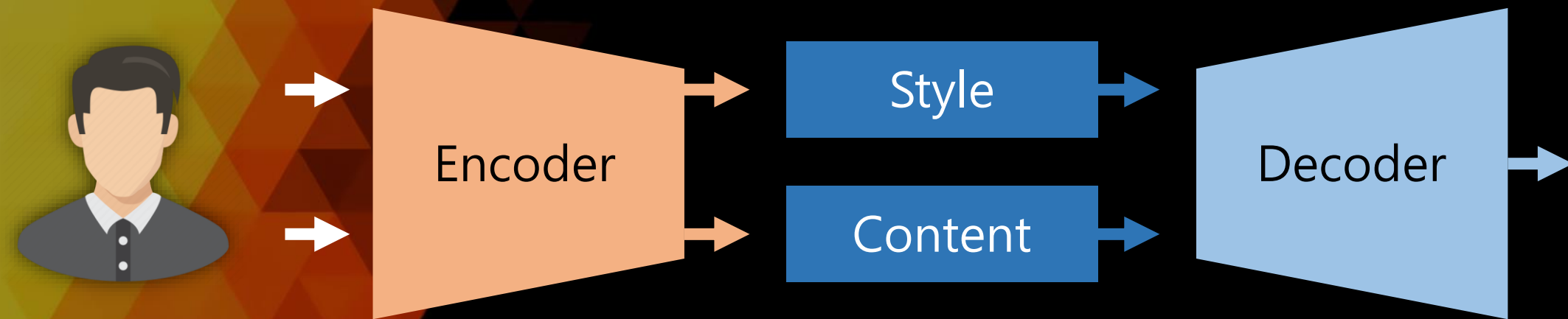
# 理論知識 – AutoVC

Εύρηκα



Εύρηκα

# 理論知識 – Training方式



Loss

元

# 訓練成果

- 聲音風格轉換



- 中文轉為英文說話的Style

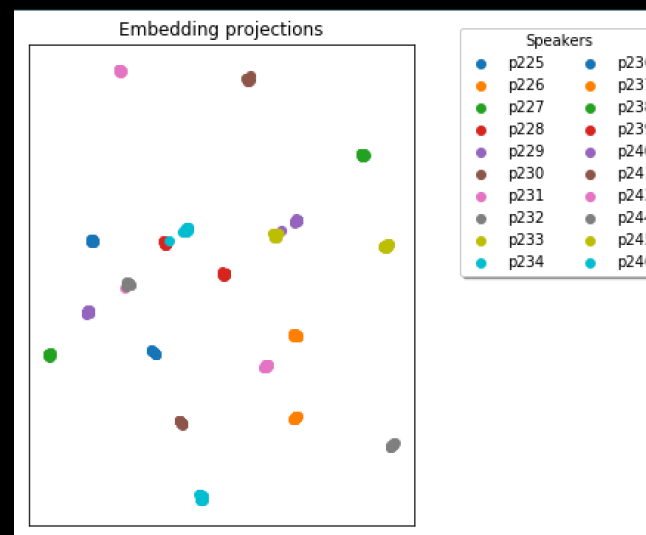
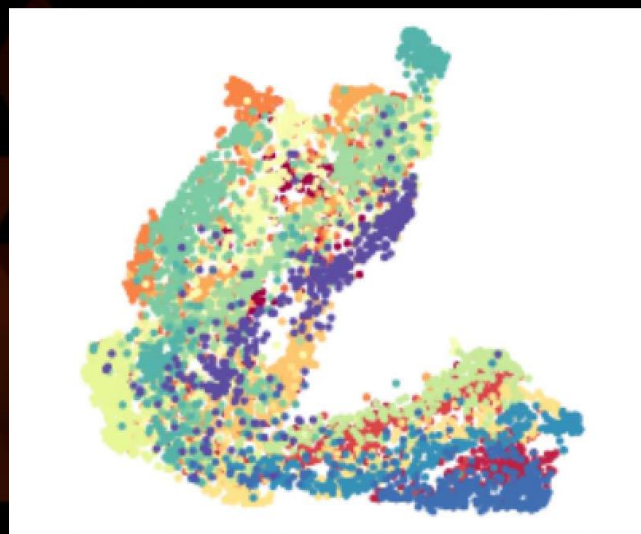


Εύρηκα

π

# 討論 – Style Encoding

Εύρηκα



One Hot

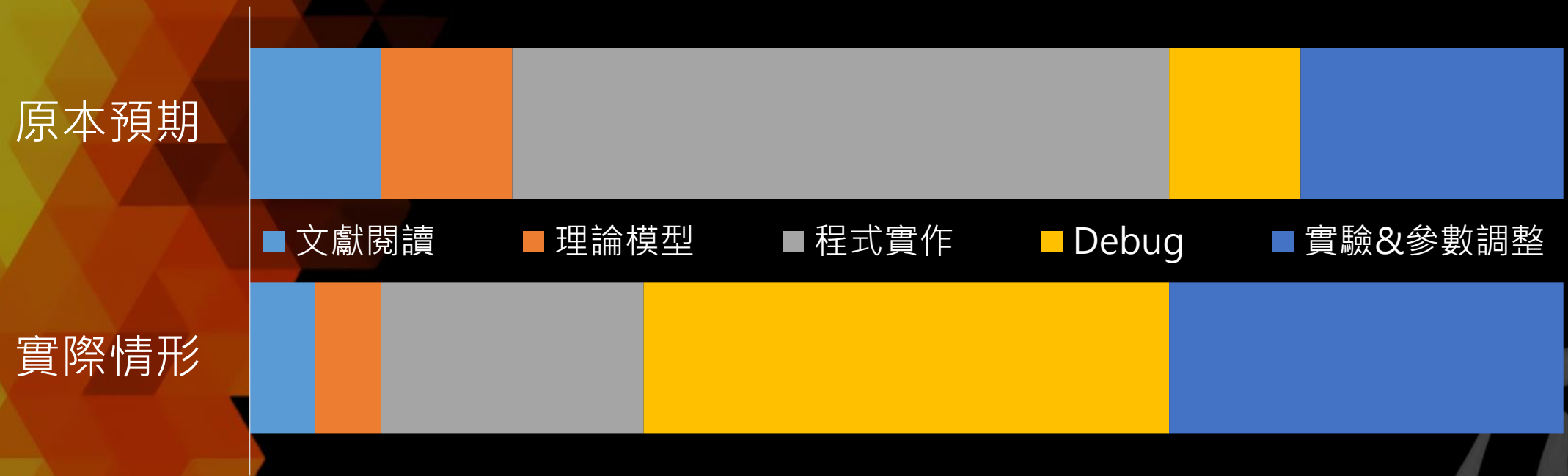
使用了錯誤的  
Encode方式

正確的Style  
Encoding

π

Εύρηκα

# 討論 – 理論 & 技術 v.s 實作



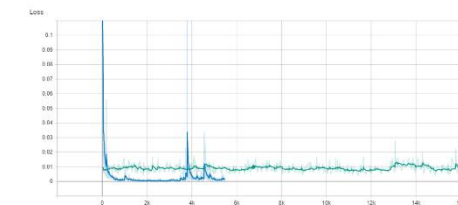
# 收穫 – 專案管理

- 程式架構的設計
- 專案架構的整理與維護
- 進度的規劃與調整
- Github連結：

[https://github.com/casperwang/autovc\\_v2](https://github.com/casperwang/autovc_v2)

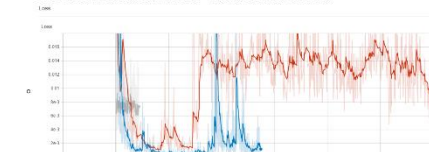
```
autovcv2
├── README.md
├── requirements.txt
├── .gitignore
├── model.py
├── train.py
├── demos
├── data
│   ├── dataloader.py
│   ├── preprocessing.py
│   └── data.pkl
├── VCTK
│   ├── p225
│   │   └── p225_001.wav
│   │   └── ...
│   └── p226
│       └── ...
├── test(inference)
│   ├── conversion.py
│   ├── vocoder.py
│   ├── test_all.py
│   └── validate.py
├── weights
│   ├── checkpoint_step001000000_ema.pth
│   └── autovc.ckpt
├── train_weights
│   └── ...
```

## 4/8 資訊專題進度報告



跑了許多次的實驗，擷取比較重要（也比較久的兩次）作報告：

- 我們找到 `train.py` 中的bug了！因為一直發現loss下降但是出來是雜訊，縱使用one-hot encoding也一樣，所以我們就再去仔細看了程式碼和paper，找到寫錯的地方（擾亂寫反qq）。
- 藍色的是用One-Hot Encoding所做出來的結果，可以看到loss急速下降，但是到約4000 iterations的時候會突然暴增，不知道為什麼。
- 此情況在其他次training的時候也有發生，就是training到一半loss突然暴增然後就慢慢掉下來或者在那邊達到水平的情况，不大知道為什麼（如下圖）





# 未來展望

- 將聲音的音質提高，與人聲更加相似
- 嘗試使用不同的Dataset訓練
- 加入影片的協助，控制五官的動作與聲音同步製造影片

Εύρηκα

π

# 參考資料

Εύρηκα

- Kaizhi Qian, et. al (2019). AUTOVC: Zero-Shot Voice Style Transfer with Only Autoencoder Loss : arXiv:1905.05879
- Quan Wang (2017). Generalized End-to-End Loss for Speaker Verification : arXiv:1710.10467
- PyTorch Library : <https://pytorch.org/>

π