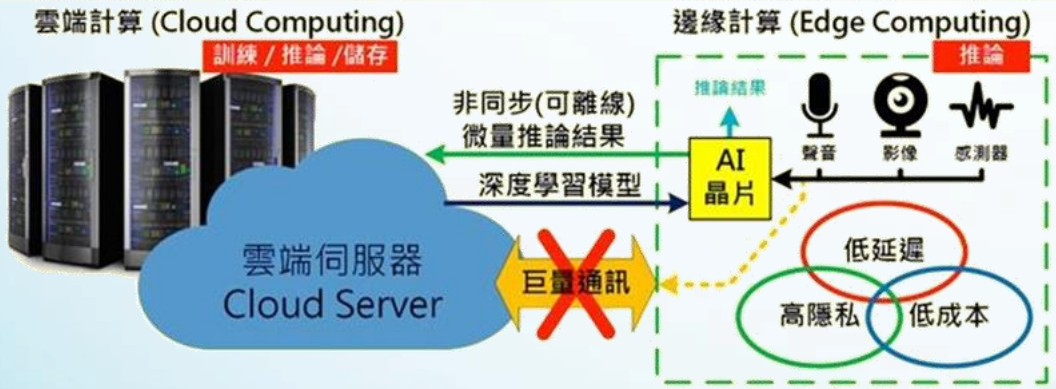
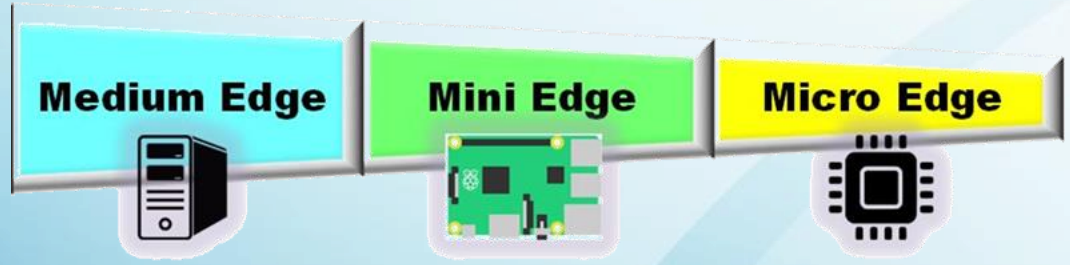
OmniXRI's Edge AI & TinyML 小學堂

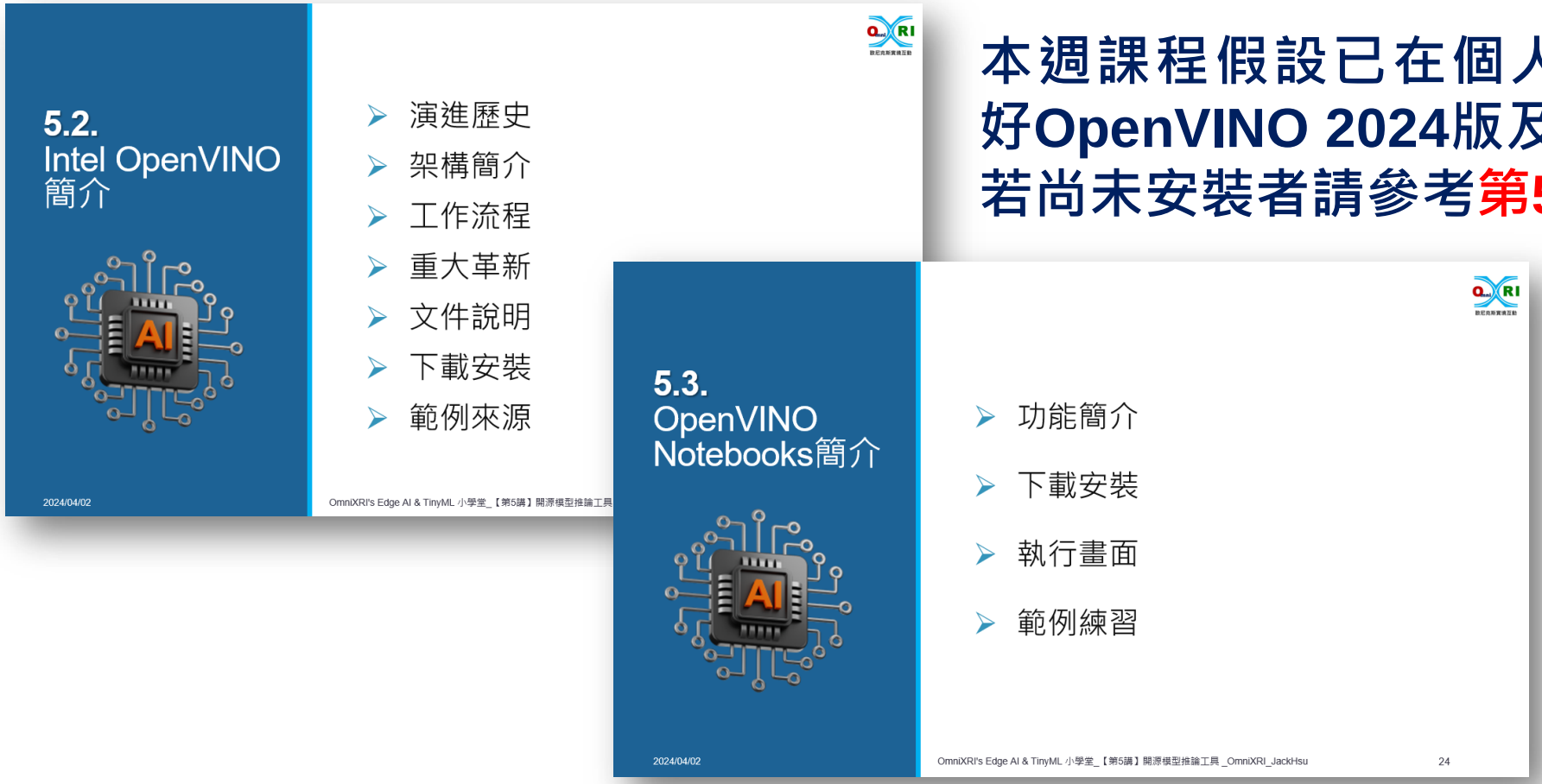




【第16講】
實作案例 – 影像音樂生成

歐尼克斯實境互動工作室 (OmniXRI Studio)
許哲豪 (Jack Hsu)

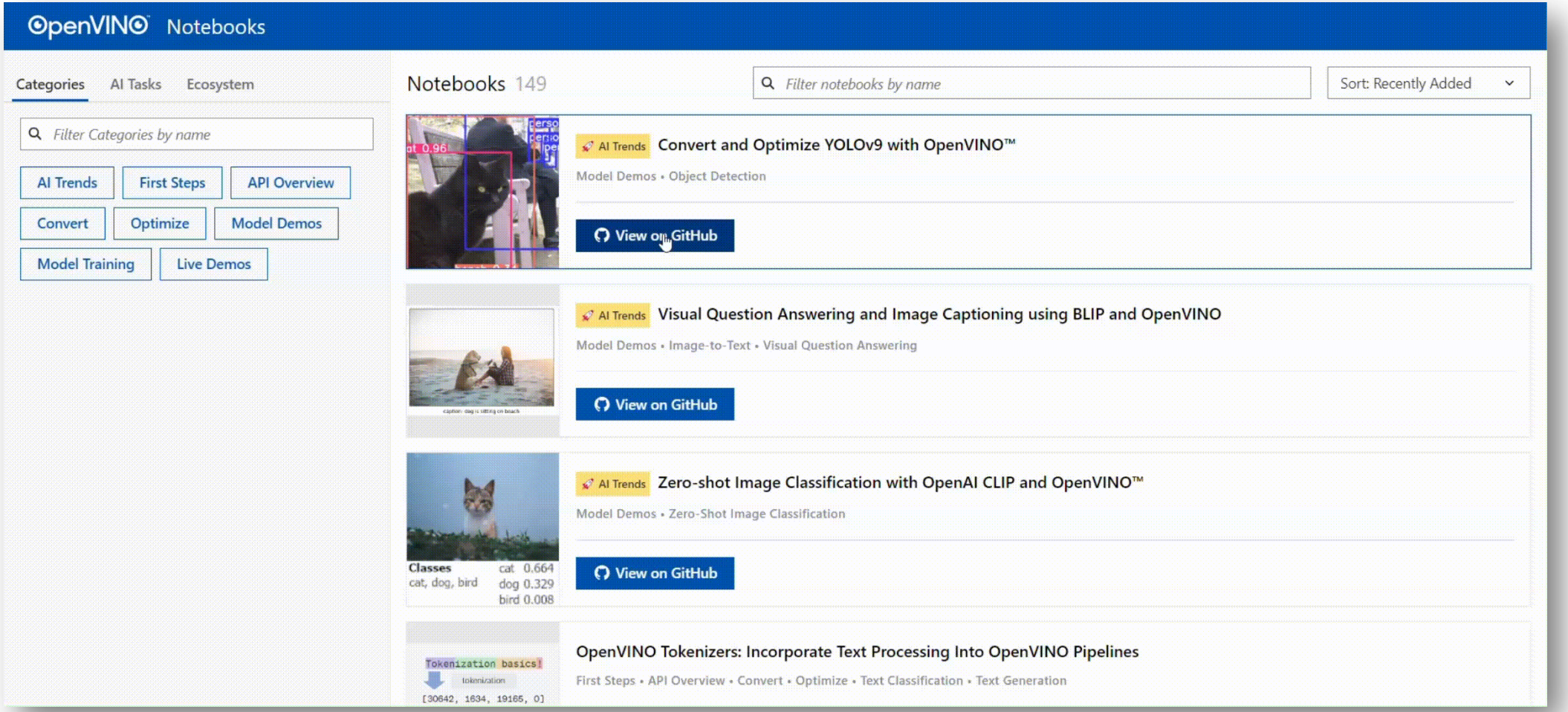
Intel OpenVINO & Notebooks 回顧



本週課程假設已在個人電腦上安裝好OpenVINO 2024版及 Notebooks。若尚未安裝者請參考第5講課程。

直播連結：https://youtu.be/6By3GXuEpFc

Intel OpenVINO Notebooks 範例程式網頁



https://openvinotoolkit.github.io/openvino_notebooks/

簡報大綱

- 16.1. 常見影像生成應用
- 16.2. 影像生成應用實例
- 16.3. 常見音樂生成應用
- 16.4. 音樂生成應用實例

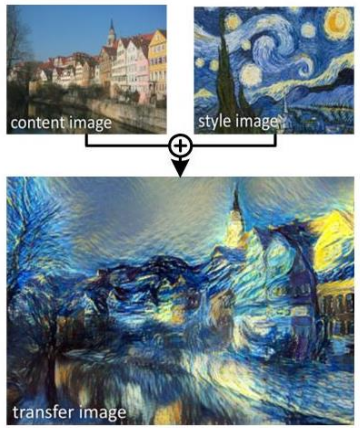
本課程完全免費，請勿移作商業用途！
歡迎留言、訂閱、點讚、轉發，讓更多需要的朋友也能一起學習。

完整課程大綱： https://omnixri.blogspot.com/2024/02/omnixris-edge-ai-tinyml-0.html
課程直播清單： https://www.youtube.com/@omnixri1784/streams

常見影像生成應用

對抗生成網路
(Generative Adversarial Network, GAN)

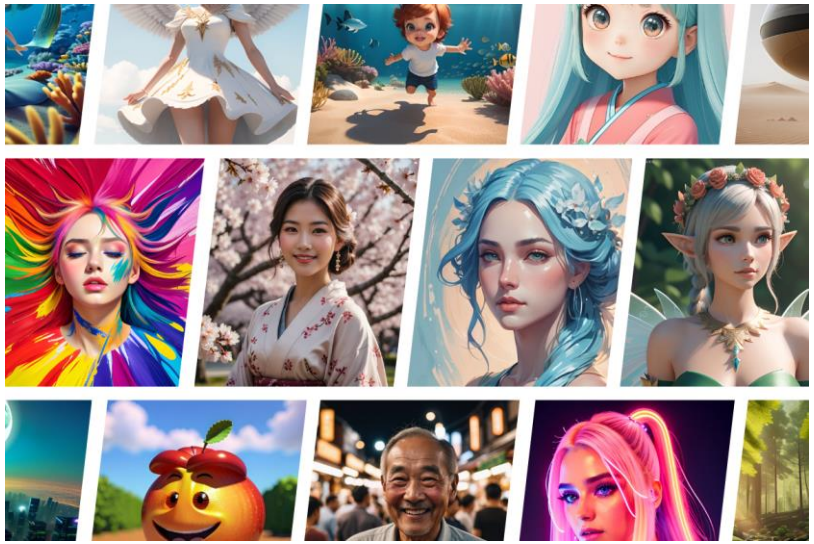
人臉生成(變臉APP)



風格轉換


圖像修復(填補/上色/去糊)

Transformer / ChatGPT生成
(AIGC, genAI)

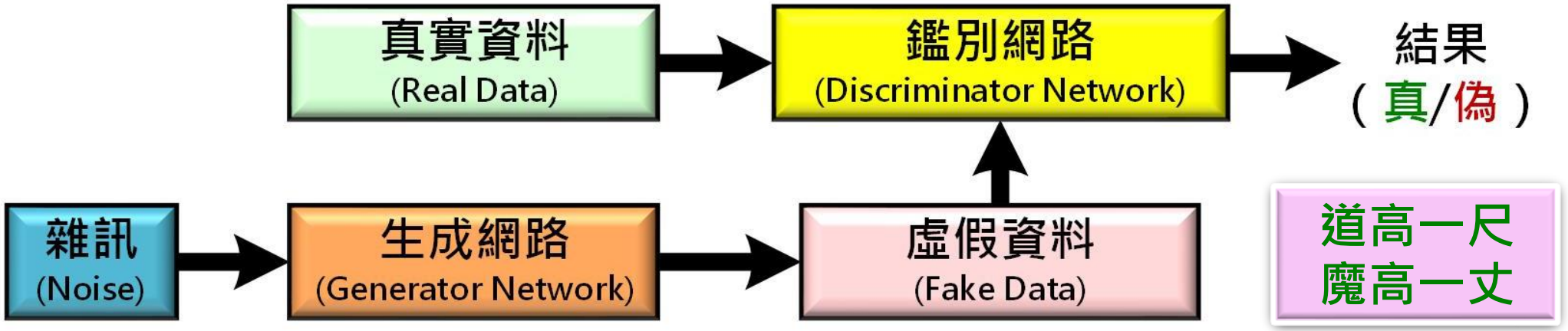


文字生成影像/影片
影像生成影像/影片
聲音生成影像/影片

- MyEdit
- Midjourney
- DALL·E3
- Stable Diffusion
- NovelAI
- Stableboost
- Fotor
- Adobe Firefly
- Leonardo.Ai
- Craiyon
- DreamStudio
- Canva

資料來源：https://tw.cyberlink.com/blog/photo-editing-tips/2345/best-ai-image-generators

影像生成 – 生成對抗網路(GAN)



人臉生成應用

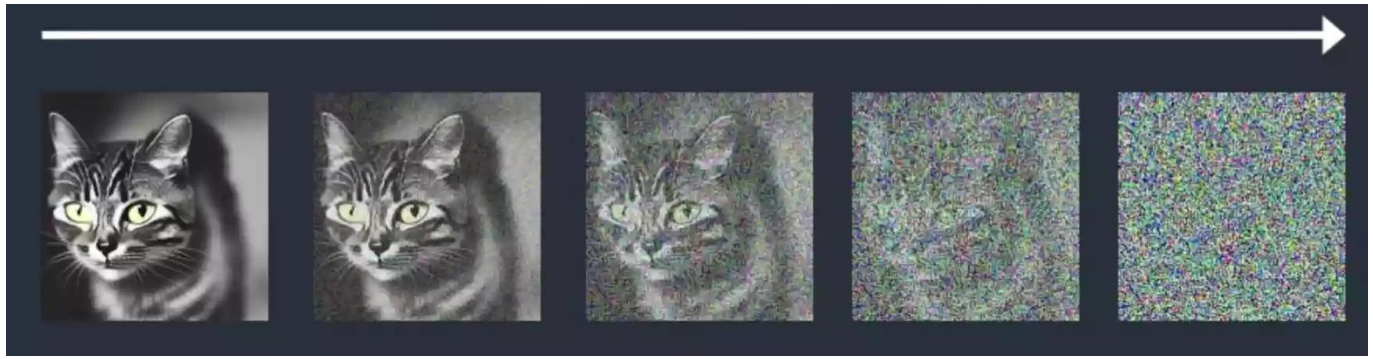
原始內容

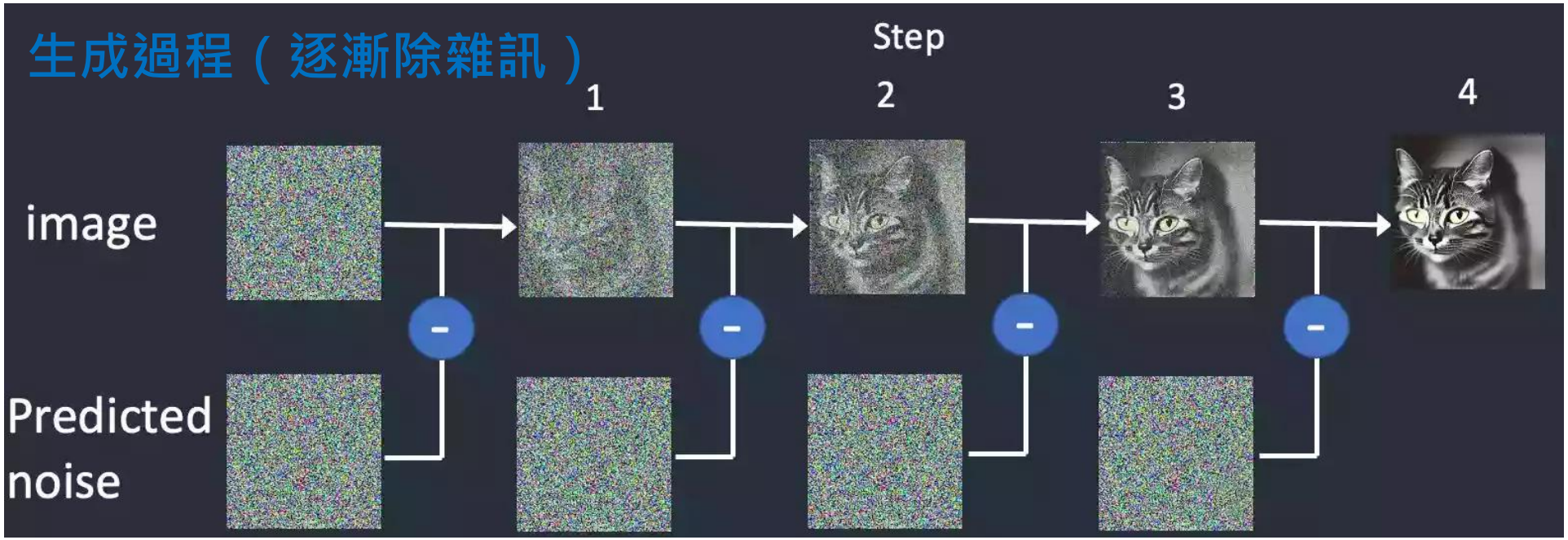
原始內容

生成內容

影像生成 – Stable Diffusion

訓練過程
（逐漸加雜訊）

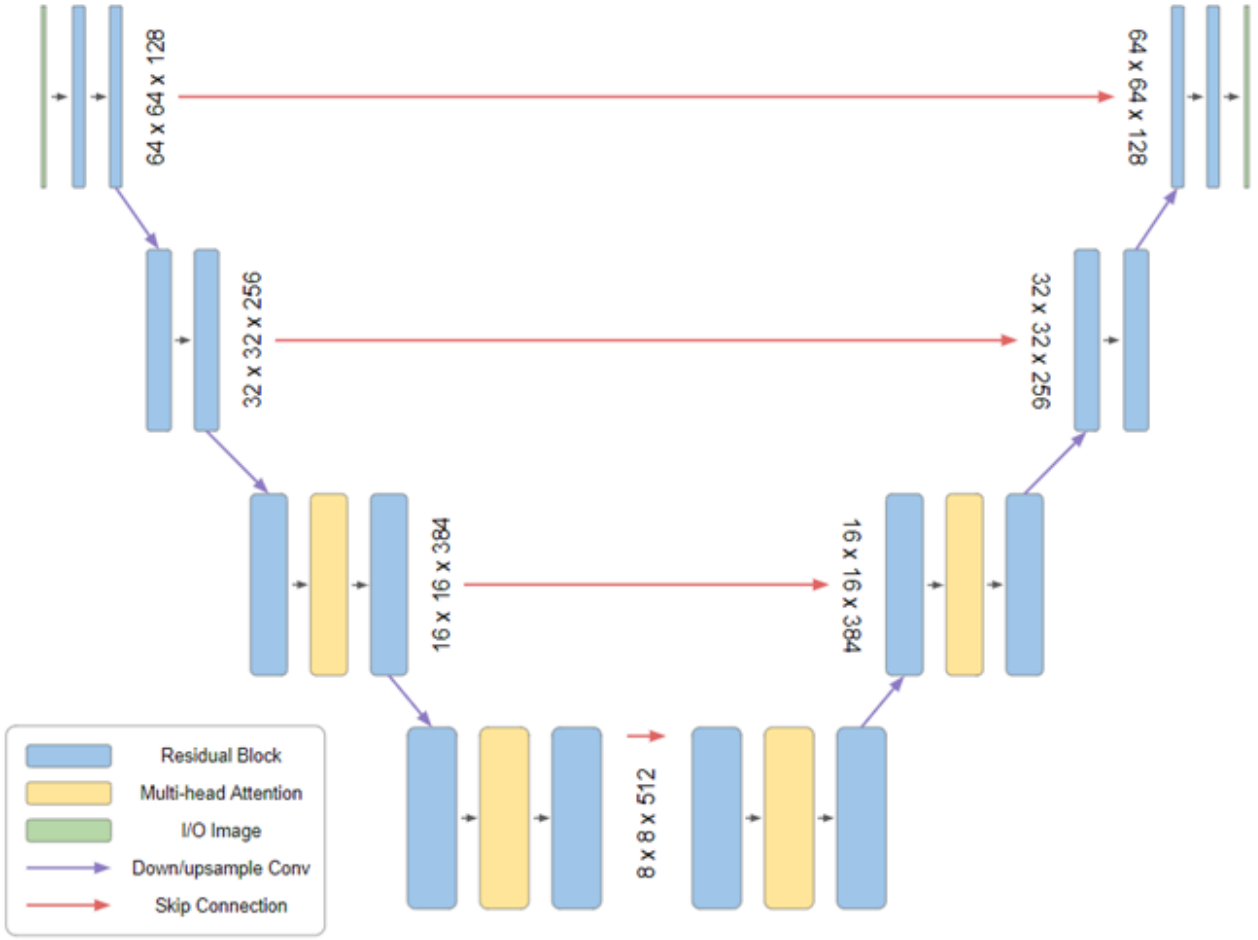




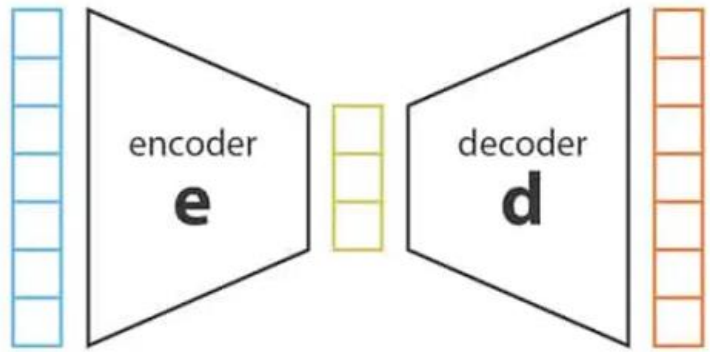
資料來源：https://openai.wiki/stable-diffusion-introduce.html

影像生成 – U-Net & VAE

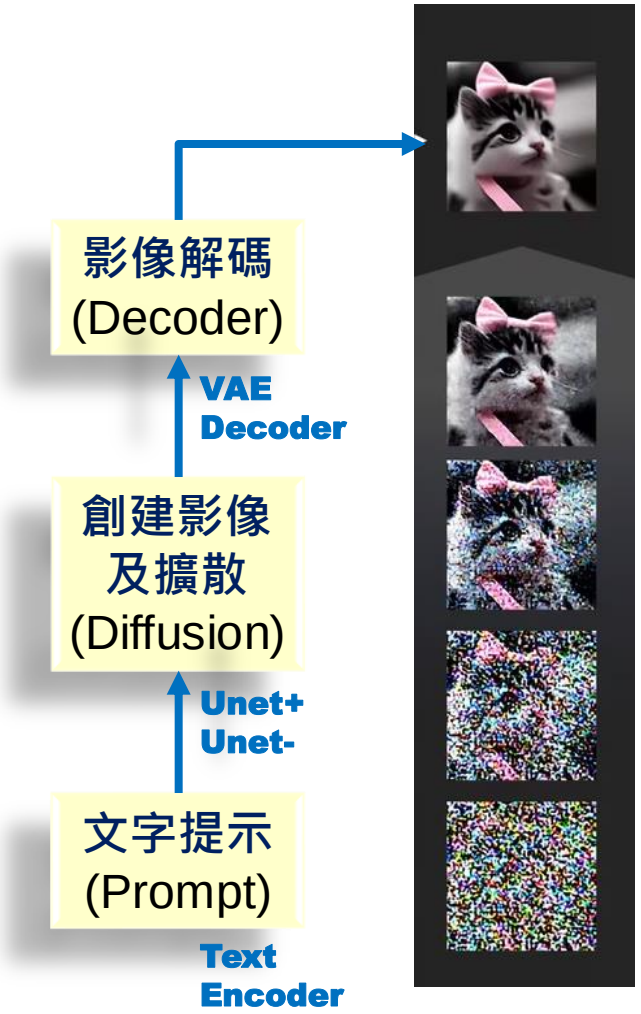
U-Net

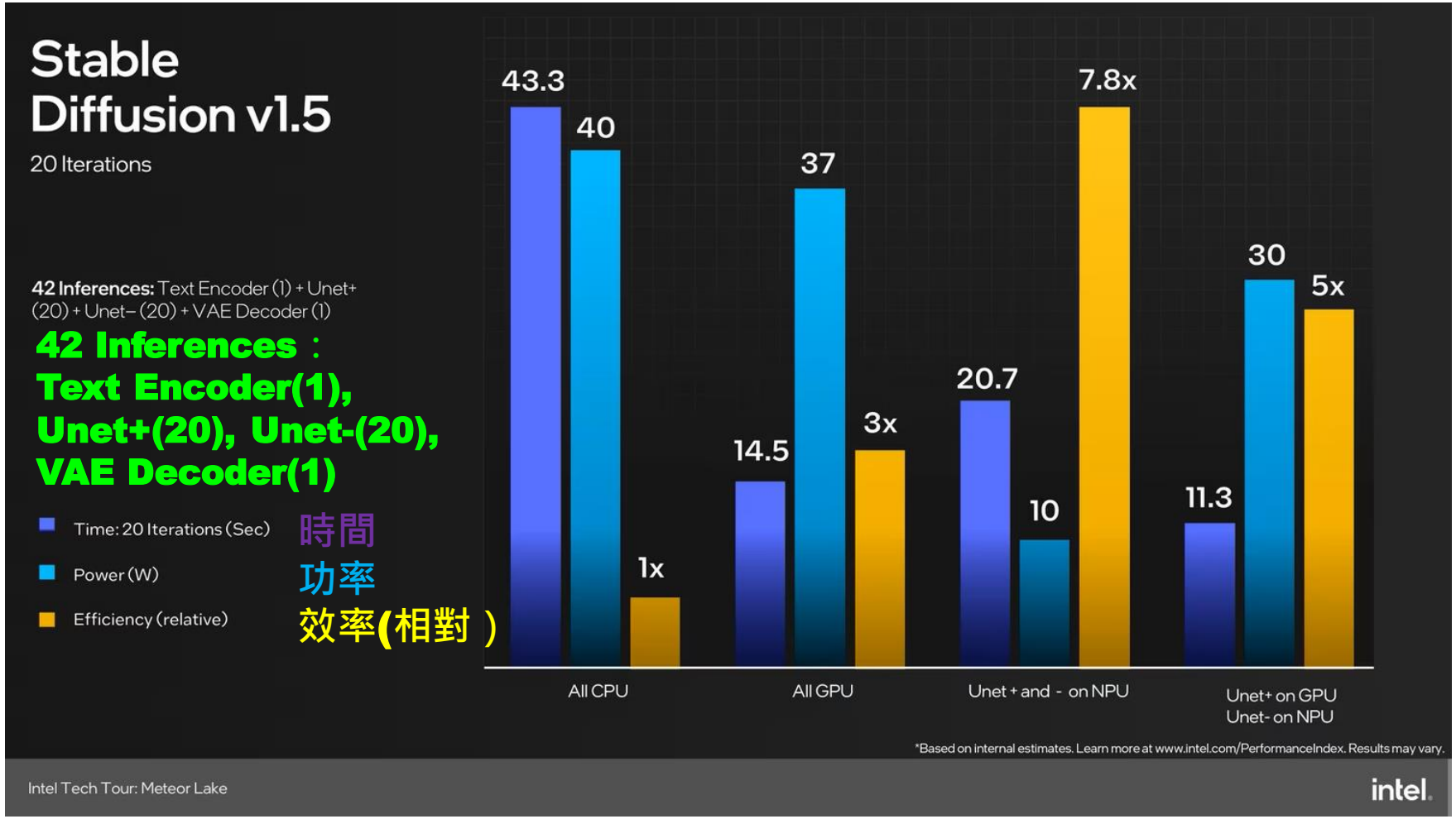


Variational Autoencoder (VAE)



Meteor Lake運行影像生成效能比較



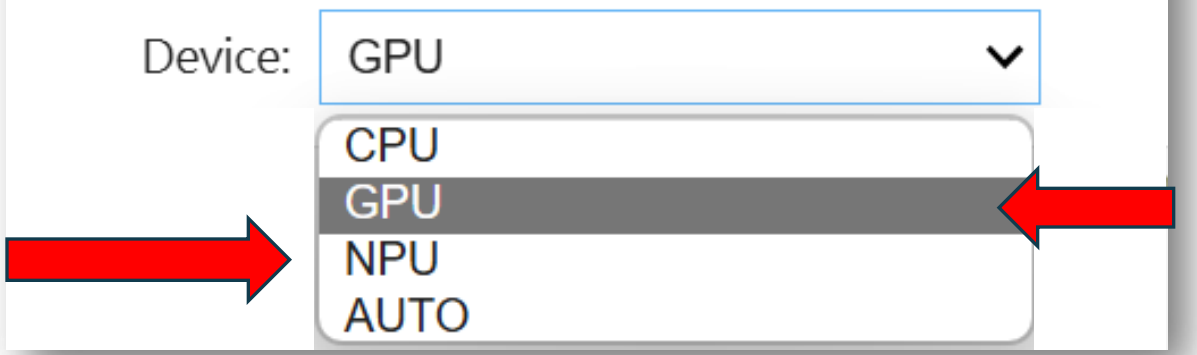


影像來源：https://www.4gamers.com.tw/news/detail/59826/intel-meteor-lake-architecture-overview

OpenVINO Notebooks Device 設定

預設為 **AUTO**
建議改為 **GPU**
以加速推論

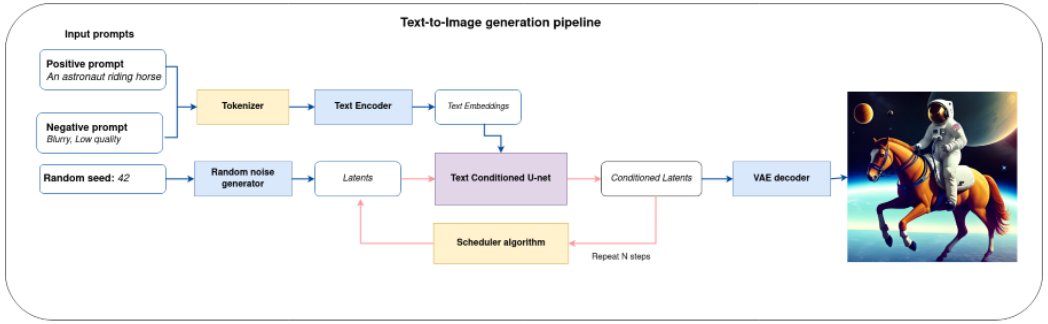
```
import ipywidgets as widgets

device = widgets.Dropdown(
    options=core.available_devices + ["AUTO"],
    value="GPU",
    description="Device:",
    disabled=False,
)

device
```



Core Ultra 才
有 **NPU** 選項

預設為內顯(**iGPU**)
12代之前為 **HD Graphic**
12代(含)之後為 **Iris Xe**

影像生成應用實例 – Tiny SD 文生圖



tiny-sd-image-generation (Colab)

使用 **Tiny SD**模型，以提示文字、種子及擴散步數來產生影像，亦可使用影像產生影像。

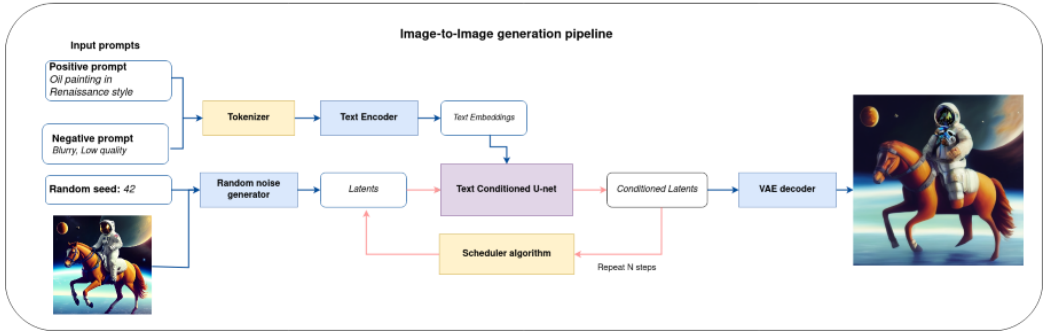
text_prompt = 'RAW studio photo of An intricate forest minitown landscape trapped in a bottle, atmospheric oliva lighting, on the table, intricate details, dark shot, soothing tones, muted colors '
seed = 431
num_steps = 20

輸出影像

- Prerequisites
- Create PyTorch Models pipeline
- Convert models to OpenVINO IR format
  - Text Encoder
  - U-net
  - VAE
- Prepare Inference Pipeline
- Configure Inference Pipeline
  - Text-to-Image generation
  - Image-to-Image generation
  - Interactive Demo

參考資料：https://colab.research.google.com/github/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/tiny-sd-image-generation/tiny-sd-image-generation.ipynb

影像生成應用實例 – Tiny SD 圖生圖



輸入影像

輸出影像

輸入一粗略的影像，可令其產生極為精細品質的影像。除一般文字提示外，亦可加入負面提示，用來抑制產生內容。

text_prompt_i2i = 'professional photo portrait of woman, highly detailed, hyper realistic, cinematic effects, soft lighting'

negative_prompt_i2i = "blurry, poor quality, low res, worst quality, cropped, ugly, poorly drawn face, without eyes, mutation, unreal, animate, poorly drawn eyes"

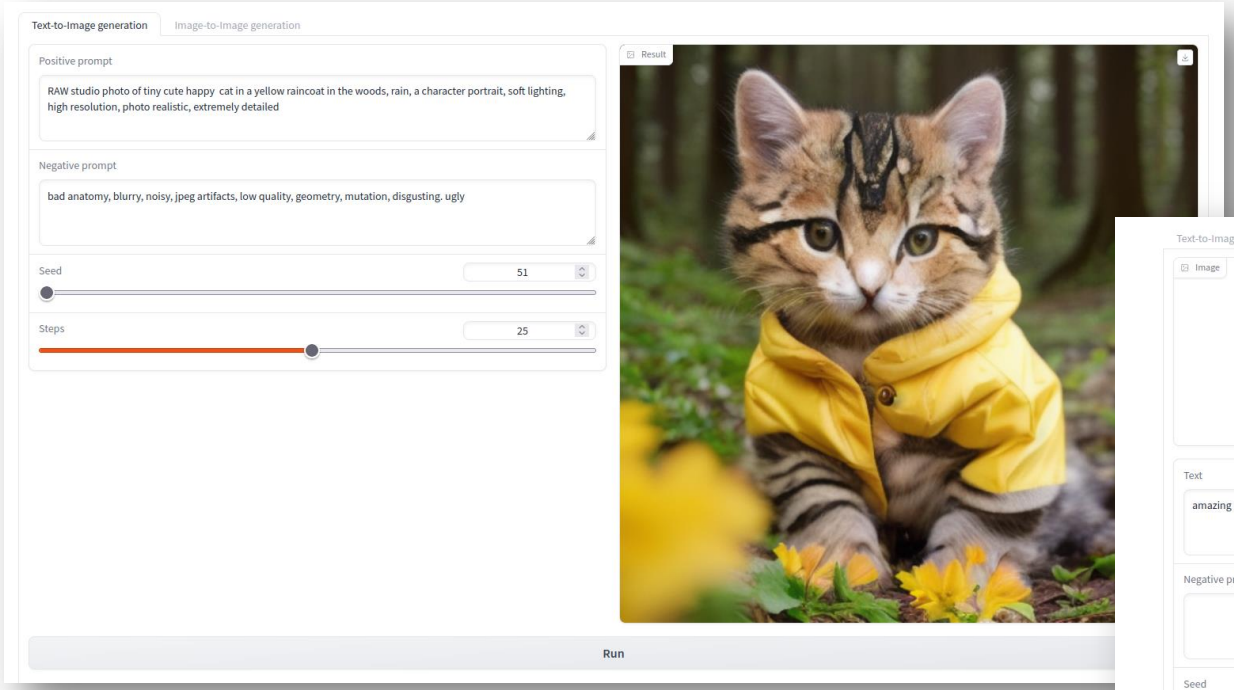
num_steps_i2i = 40

seed_i2i = 82698152

strength = 0.68

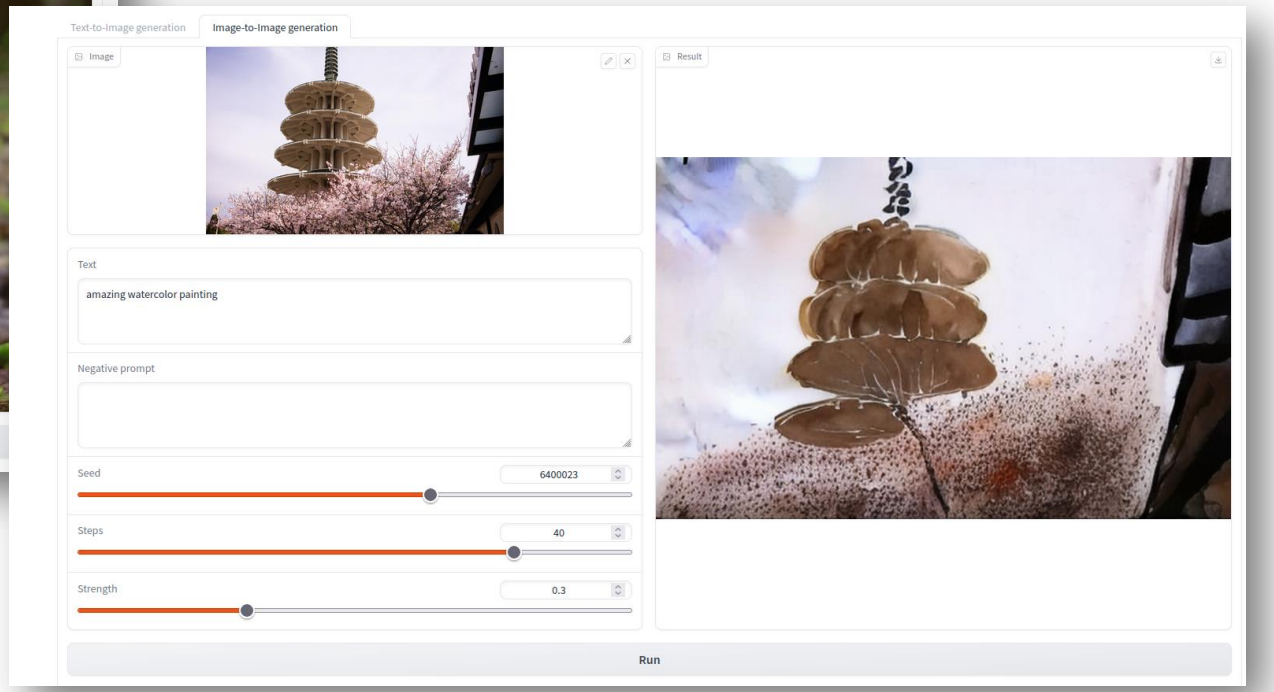
參考資料：https://colab.research.google.com/github/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/tiny-sd-image-generation/tiny-sd-image-generation.ipynb

影像生成應用實例 – Tiny SD 互動介面



文字生成影像

互動式操作介面



影像生成影像

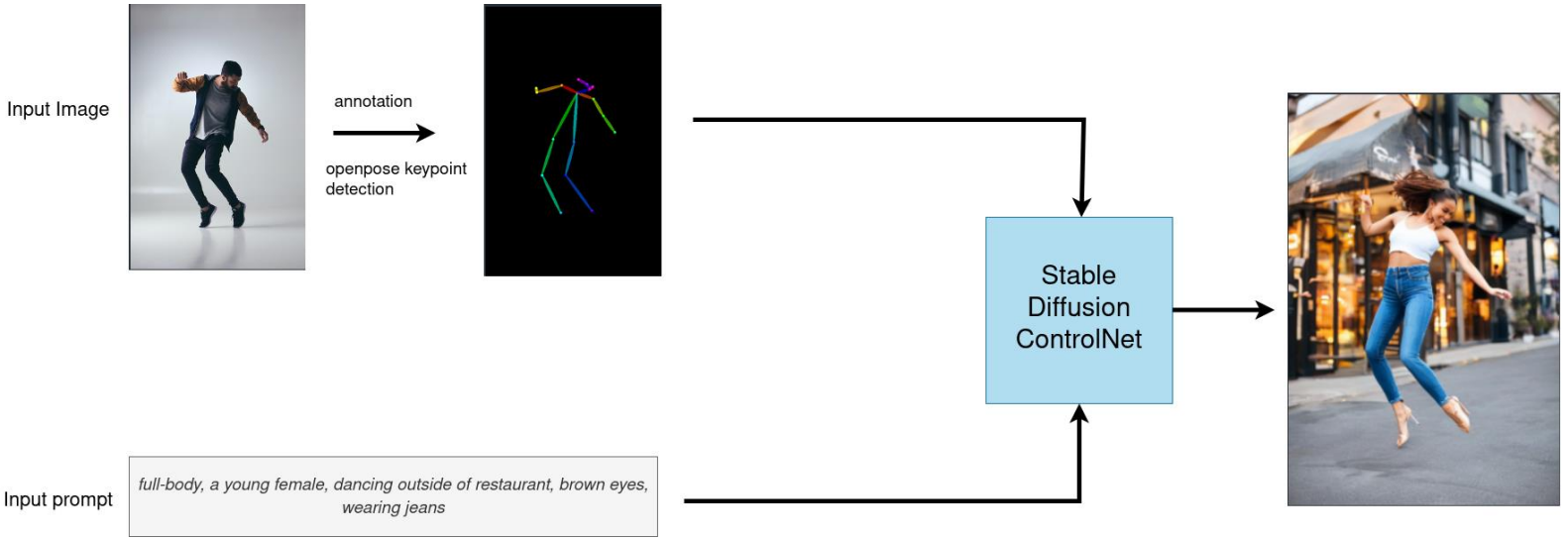
參考資料：https://colab.research.google.com/github/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/tiny-sd-image-generation/tiny-sd-image-generation.ipynb

影像生成應用實例 – ControlNet 圖生圖

Text-to-Image Generation with ControlNet Conditioning

controlnet-stable-diffusion.ipynb

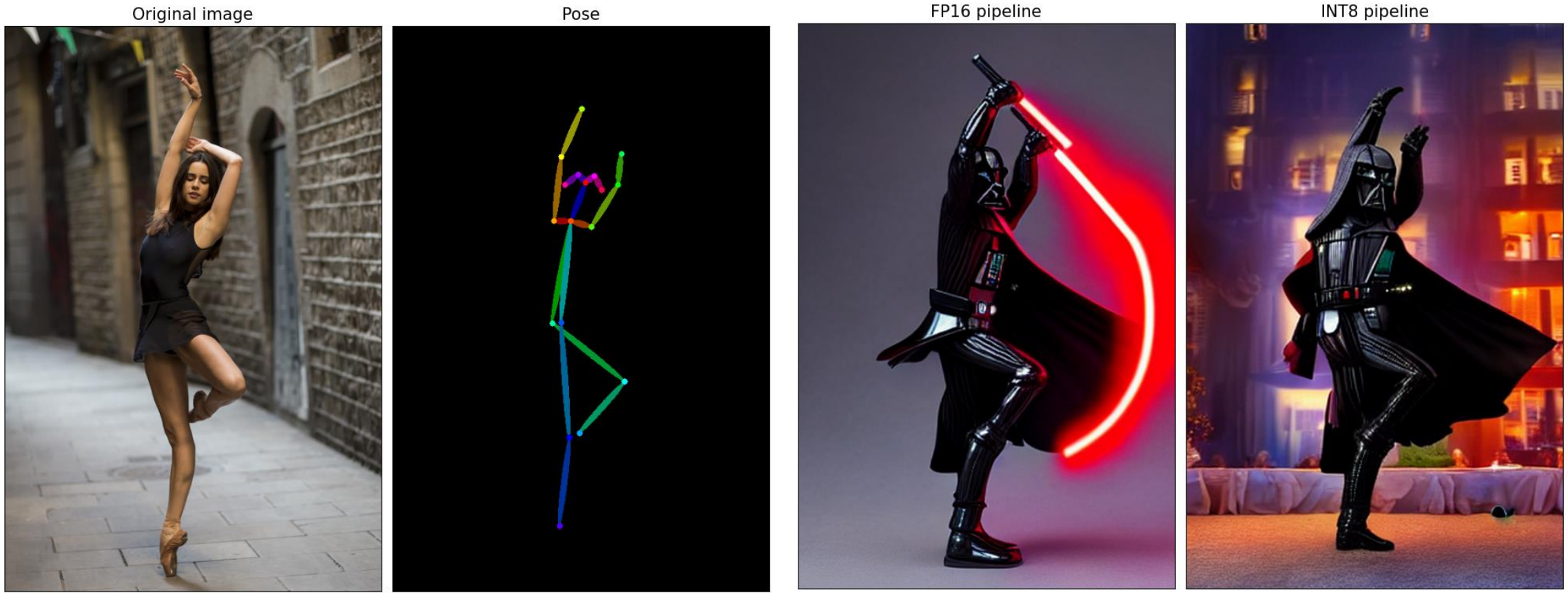
將影像中人物姿態以**OpenPose**取出，再搭配**輸入文字**及 **ControlNet** 模型產生新影像。



- Instantiating Generation Pipeline
  - **ControlNet in Diffusers library**
  - **OpenPose**
- Convert models to OpenVINO Intermediate representation (IR) format-format)
  - **OpenPose conversion**
- Select inference device
  - **ControlNet conversion**
  - **UNet conversion**
  - **Text Encoder**
  - **VAE Decoder conversion**
- Prepare Inference pipeline
- Running Text-to-Image Generation with ControlNet Conditioning and OpenVINO
- Select inference device for Stable Diffusion pipeline
- Quantization
- Interactive demo

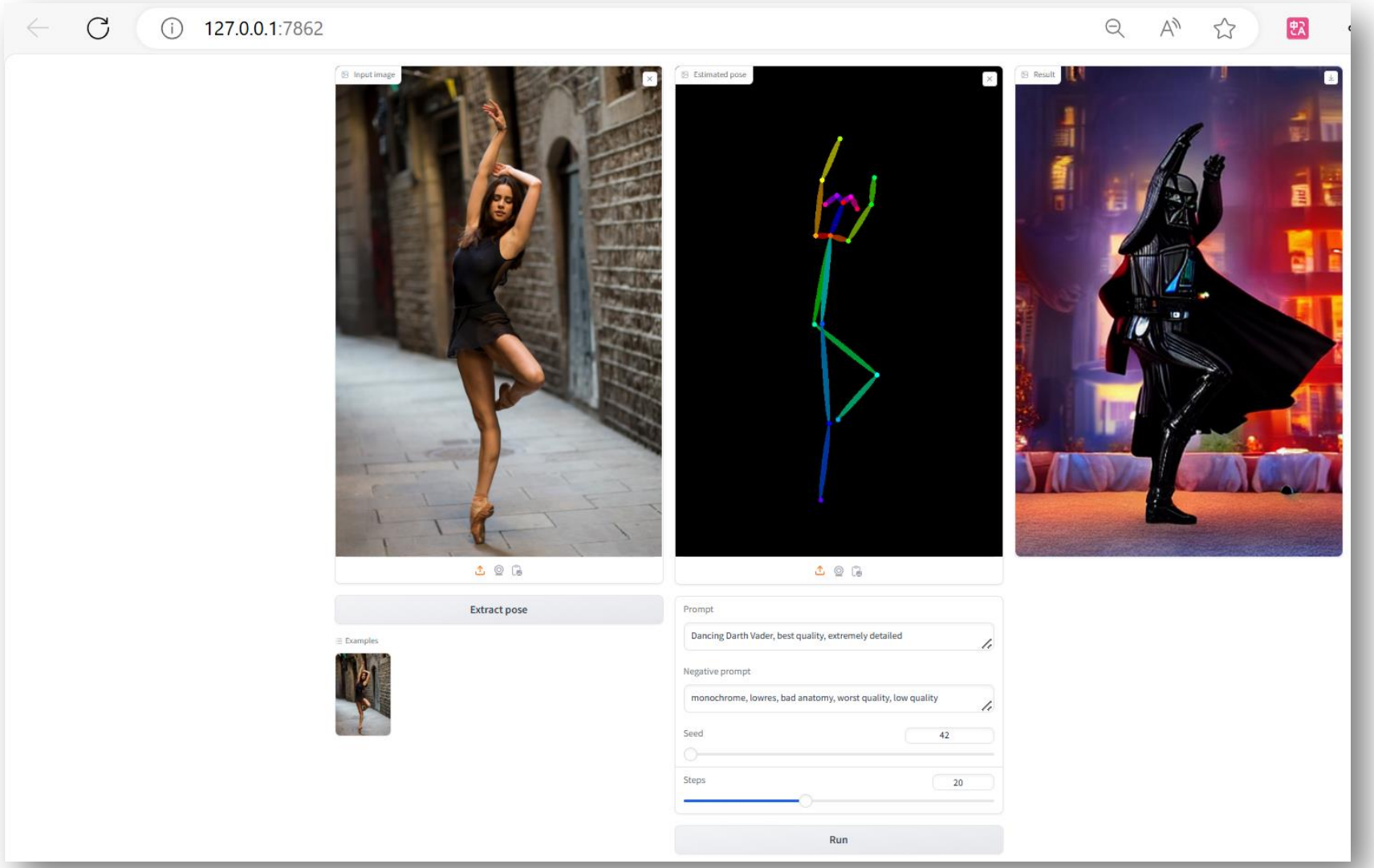
資料來源：https://github.com/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/controlnet-stable-diffusion/controlnet-stable-diffusion.ipynb

影像生成應用實例 – ControlNet 結果




```
prompt = "Dancing Darth Vader, best quality, extremely detailed"
negative_prompt = "monochrome, lowres, bad anatomy, worst quality, low quality"
result = ov_pipe(prompt, pose, 20, negative_prompt=negative_prompt)
```

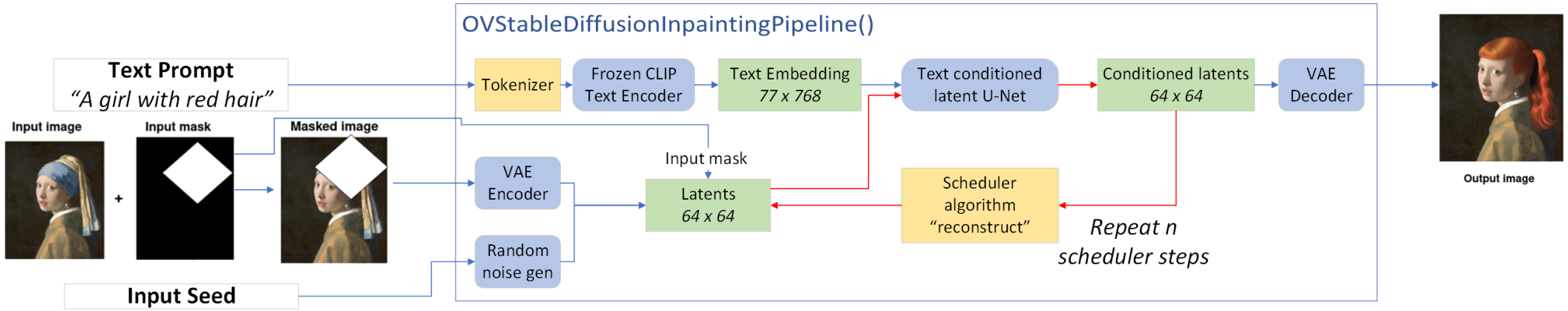
影像生成應用實例 – ControlNet 互動介面



影像生成應用實例 – Infinite Zoom

Infinite Zoom Stable Diffusion v2 and OpenVINO

\stable-diffusion-v2\stable-diffusion-v2-infinite-zoom.ipynb



- Stable Diffusion v2 Infinite Zoom Showcase
  - Stable Diffusion Text guided Inpainting
- Prerequisites
  - Stable Diffusion in Diffusers library
  - Convert models to OpenVINO Intermediate representation (IR) format-format)

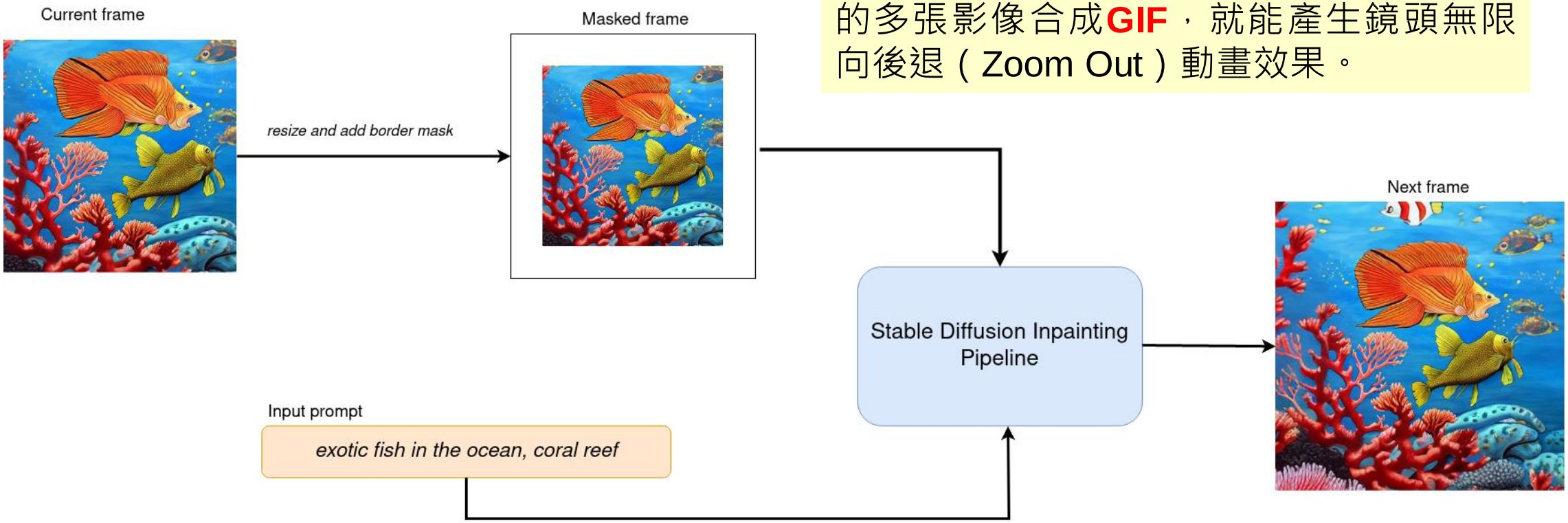
將影像**缺失**部份加上**文字輸入**再使用 **Stable Diffusion v2** 模型產生新影像補齊缺失部份。

- Prepare Inference pipeline
- Zoom Video Generation
- Configure Inference Pipeline
- Select inference device
- Run Infinite Zoom video generation

資料來源：https://github.com/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/stable-diffusion-v2/stable-diffusion-v2-infinite-zoom.ipynb

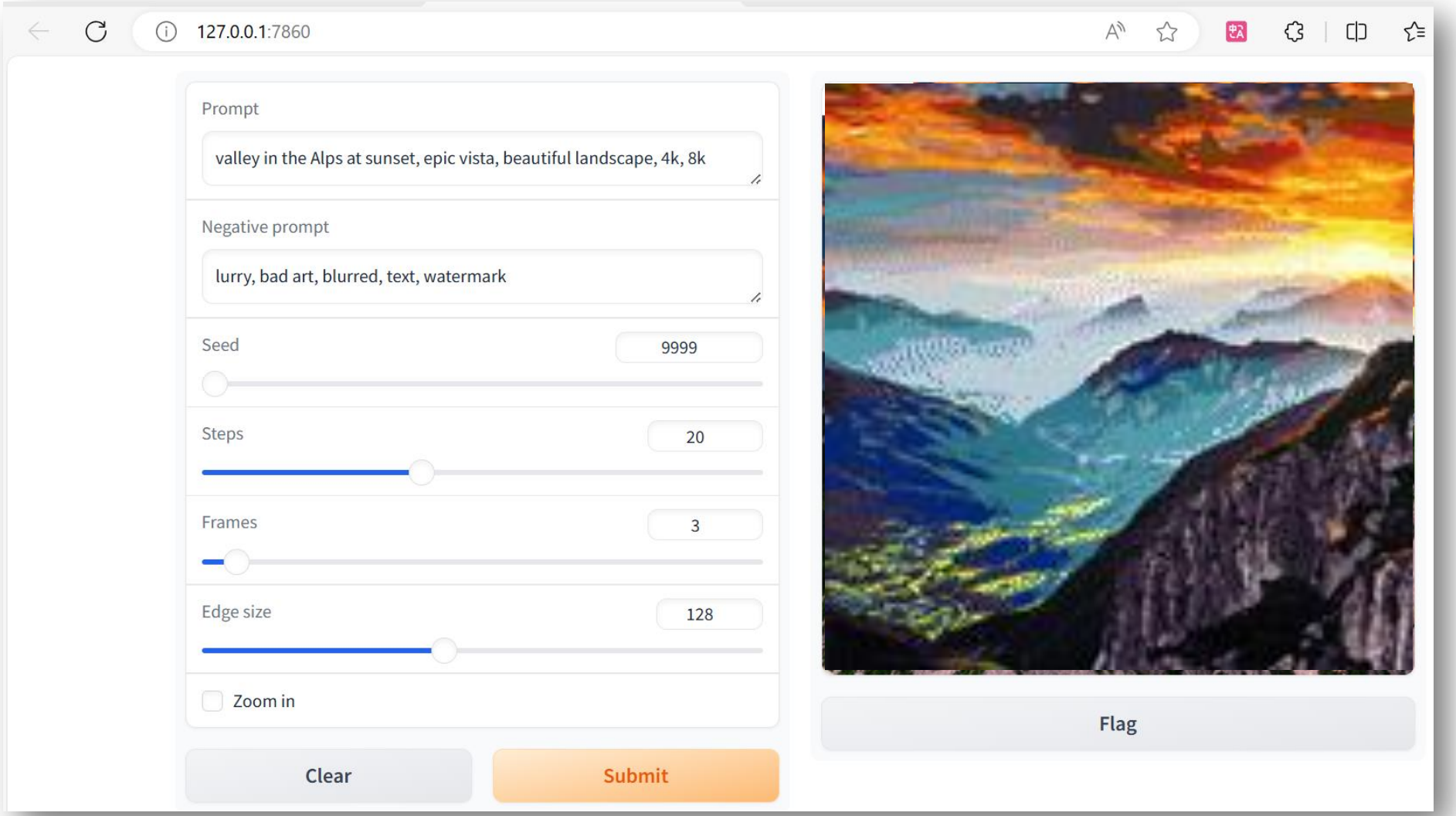
影像生成應用實例 – Infinite Zoom 工作流程

將原始影像縮小令外圍產生一圈空白，加上**文字輸入**，再使用 **Stable Diffusion v2** 模型產生新影像補齊缺失部份。再將完成的多張影像合成**GIF**，就能產生鏡頭無限向後退（Zoom Out）動畫效果。

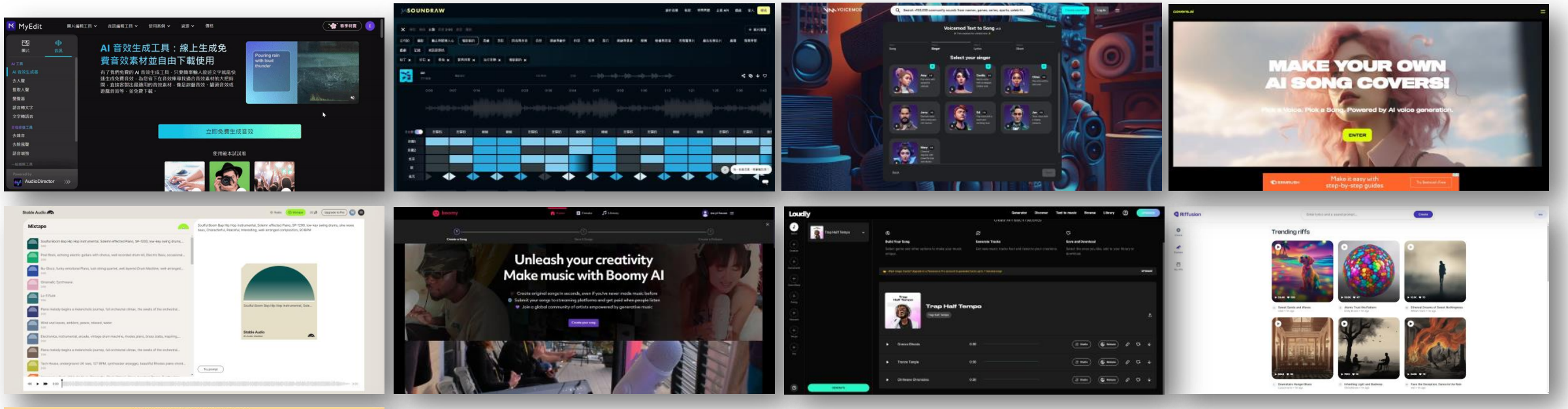


資料來源：https://github.com/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/stable-diffusion-v2/stable-diffusion-v2-infinite-zoom.ipynb

影像生成應用實例 – Infinite Zoom 互動介面



常見音樂生成應用

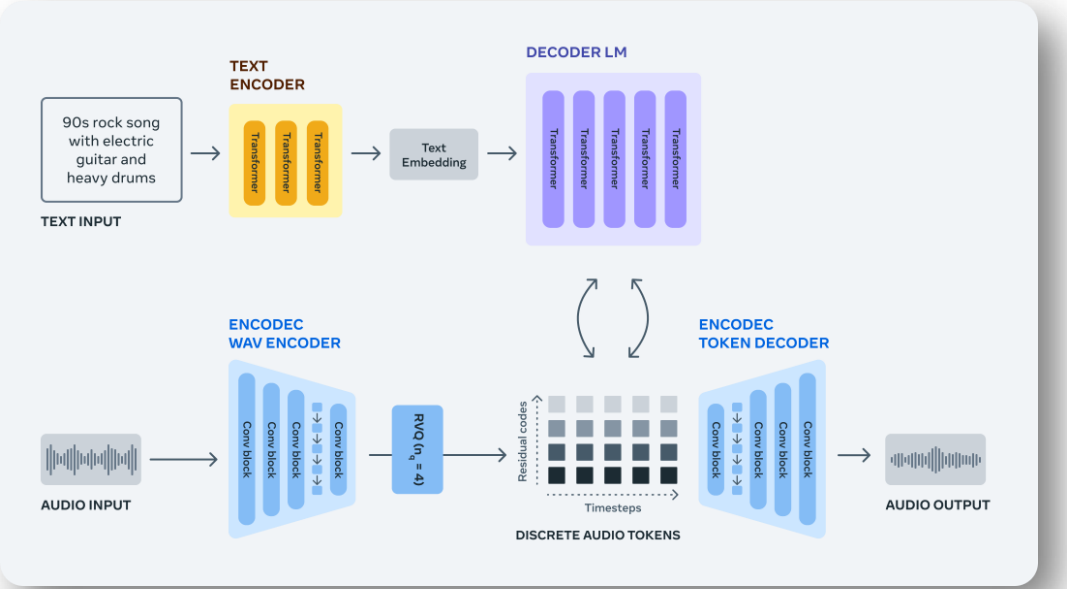


1. [MyEdit：免費線上 AI 音效生成器]
2. [Stable Audio： 線上生成高音質純音樂]
3. [Suno AI：AI 生成有人聲的歌曲]
4. [Soundraw：AI 生成主題配樂]
5. [Boomy：AI 音樂經紀人]
6. [Riffusion：AI 唱歌工具]
7. [Voicemod ：可指定虛擬歌手]
8. [Loudly：豐富音樂客製化功能]
9. [covers.ai：AI 翻唱工具]

資料來源：https://tw.cyberlink.com/blog/the-top-audio-editors/3014/online-ai-music-generator

音樂生成應用實例 – MusicGen 字生樂

```
inputs = processor(
    text=["80s pop track with bassy drums and synth"],
    return_tensors="pt",
)
```



music-generation

使用 **MusicGen** 模型，以文字提示用以產生音樂片段。

工作流程：

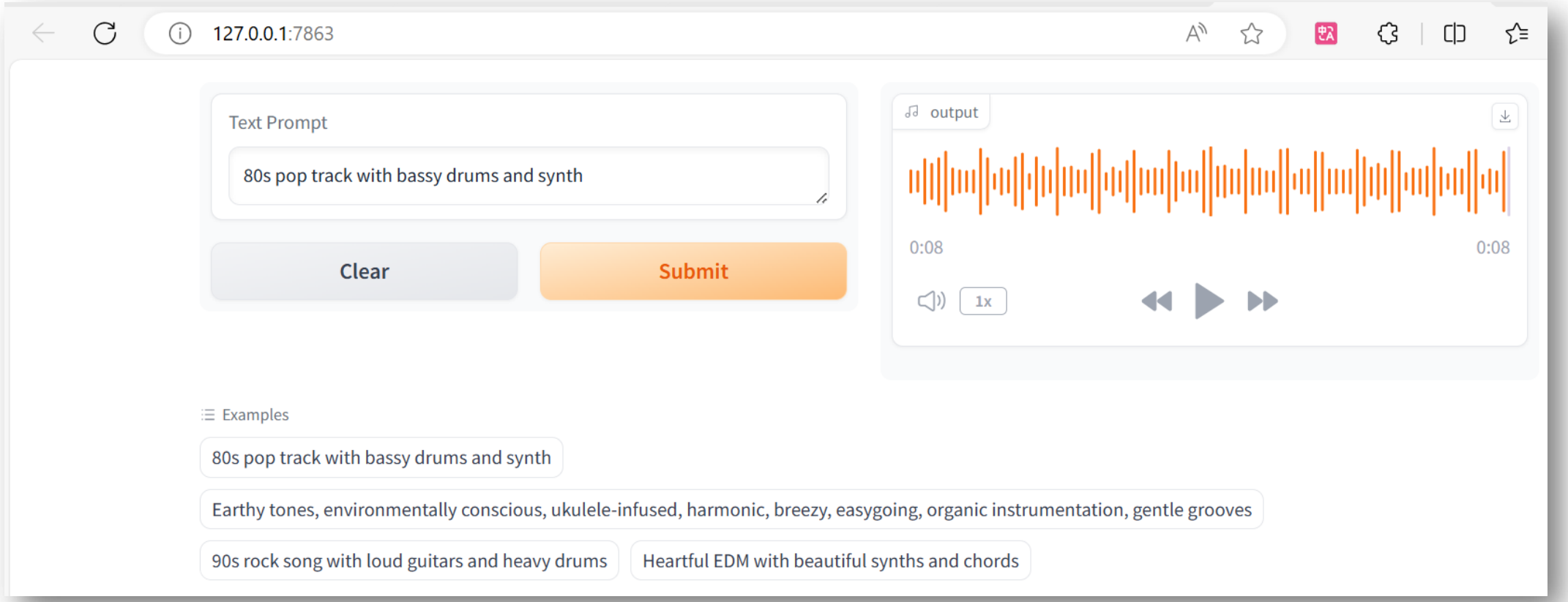
- Prerequisites
  - Install requirements
  - Imports
- MusicGen in HF Transformers
  - Original Pipeline Inference
- Convert models to OpenVINO Intermediate representation (IR) format
  - 0. Set Up Variables
  - 1. Convert Text Encoder
  - 2. Convert MusicGen Language Model
  - 3. Convert Audio Decoder
- Embedding the converted models into the original pipeline
  - Select inference device
  - Adapt OpenVINO models to the original pipeline
- Try out the converted pipeline

參考資料：https://colab.research.google.com/github/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/music-generation/music-generation.ipynb

音樂生成應用實例 – MusicGen 互動介面

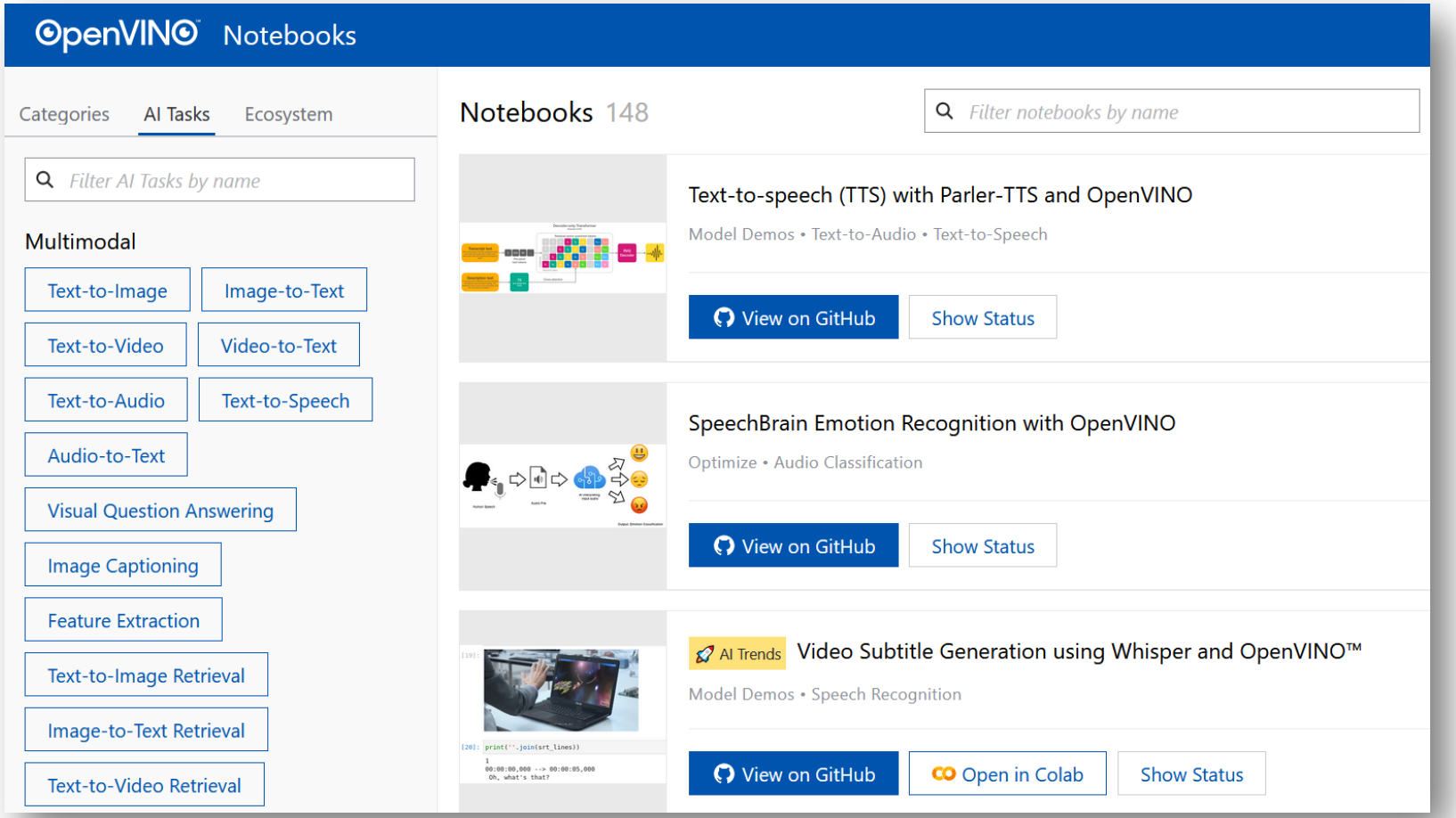
互動式操作介面

輸出結果：



參考資料：https://github.com/openvinotoolkit/openvino_notebooks/tree/main/notebooks/250-music-generation

更多 Intel OpenVINO AIGC 範例



https://openvinotoolkit.github.io/openvino_notebooks/

參考文獻

- 許哲豪，臺灣科技大學資訊工程系「人工智慧與邊緣運算實務」（2021~2023）

https://omnixri.blogspot.com/p/ntust-edge-ai.html

- 許哲豪，【課程簡報】20231209_DevFest Taichung_如何結合Google Colab及Intel OpenVINO來玩轉AIGC

https://omnixri.blogspot.com/2023/12/20231209devfest-taichunggoogle.html

延伸閱讀

- Intel OpenVINO DevCon（Youtube 中文講座）

https://www.youtube.com/watch?v=jnYNJVvghgE&list=PLJhgRo1wc4K9LRAUUgG-48BxJVqXEXhhH

- Intel OpenVINO™ 生成式 AI 系列 (Bilibili 教學影片)

https://space.bilibili.com/38566875/channel/collectiondetail?sid=2301246

感謝大家這 16 週的支持

完整課程影片：https://www.youtube.com/@omnixri1784/streams
完整課程簡報：https://github.com/OmniXRI/Edge_AI_TinyML_Course_2024

歐尼克斯實境互動工作室
(OmniXRI Studio)
許哲豪 (Jack Hsu)

Facebook : Jack Omnixri
FB社團： Edge AI Taiwan邊緣智能交流區
電子信箱：omnixri@gmail.com
部落格：https://omnixri.blogspot.tw
開　源：https://github.com/OmniXRI

YOUTUBE 直播： https://www.youtube.com/@omnixri1784/streams