

經費來源：☐01 公務 ☒02 非公務

機密(E)：☐是 ☒否

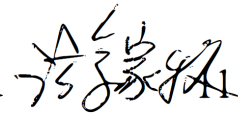
分項計畫名稱：

財團法人國家實驗研究院
資安研究能量整合計畫
結案報告書

編 號：

成果報告適用期間： 110.02.01~110.05.31

研 究 生 姓 名： 周信廷

指 導 教 授： 游家牧  111 年 5 月 20 日

計畫共同主持人： 游家牧  111 年 5 月 20 日

計 畫 主 持 人： 111 年 月 日

一、基本資料

| | | | | | | | | | | |
|-----------------------|---|--------------|---|------|-------|---|---|----|----|--|
| 計畫名稱 | 財團法人國家實驗研究院資安研究能量整合計畫 | | | | | | | | | |
| 研究生 | 周信廷 | | | | | | | | | |
| 指導教授 | 游家牧 | | | | | | | | | |
| 就讀學校 | 學校： | 國立陽明交通大學 | | | | | | | | |
| | 科系： | 財務金融研究所資料科學組 | | | | | | | | |
| | 年級： | 一 | | | | | | | | |
| 研究題目 | Deepfake 偵測分析技術 | | | | | | | | | |
| 研究或相關學習事項摘要(50-200字內) | <p>1. 建立深度學習模型分析各類 Deepfake 影片，學習影片中之換臉特徵，用於偵測影像或監視錄影是否受到 Deepfake 技術之攻擊。</p> <p>2. 將建立之深度學習模型部署至自行架設之網站，以提供社會大眾使用。</p> | | | | | | | | | |
| 每月獎助金 (研究津貼) | <input type="checkbox"/> 博士 <input checked="" type="checkbox"/> 碩士 | 10,000 | 元 | 執行期間 | 111 年 | 2 | 月 | 1 | 日 | |
| | | | | | 至 | | | | | |
| | | | | | 111 年 | 5 | 月 | 31 | 日止 | |

二、研究主題：Deepfake 偵測分析技術

三、研究問題及先前研究調查比較(Survey)：

深度臉部偽造技術（Deepfake）是一種將目標人臉移植到影片中的原始人臉，造成侵犯版權、資訊混淆甚至造成公眾恐慌等嚴重問題的惡意技術。由於人臉包含了豐富的個人資訊，濫用 Deepfake 將成為一種威脅。

最近烏俄戰爭流出烏克蘭總統澤倫斯基要求烏軍投降的 Deepfake 換臉影片在網路上瘋傳；臺灣 YouTuber 小玉透過 Deepfake 將公眾人物的臉移花接木成色情影片牟利，嚴重貶損當事人形象，震驚社會；美國前總統歐巴馬公開辱罵川普的偽造演講，也在網路上引起了極大的關注。除了名人，由於社交平台上大量的影片片段和可以免費獲取的 Deepfake 技術，使普通人也可能成為 Deepfake 的受害者。因此，如何檢測 Deepfake 影片成為了當務之急。

自從 Deepfake 日益進步並引起嚴重的社會問題以及國家安全問題，學者們積極發展深度臉部偽造檢測技術（Deepfake Detection）以對抗 Deepfake。

到目前為止，Deepfake 的檢測方法大致可以分為兩種。第一種主要關注影片單一幀中的缺陷與破綻。第二種考慮了時間相關特徵。然而，有一些方法主要針對 Deepfake 技術的非本質缺陷，例如異常眨眼或不同顏色的虹膜，反過來又刺激了 Deepfake 影片合成的進步。所以現在成為了 Deepfake 與 Deepfake Detection 的技術角力戰。

四、研究方法及步驟

本研究之目的為建立深度學習模型進行 Deepfake Detection，分析各類 Deepfake 影片，並根據各種不同的 Deepfake 方法，學習相對應的換臉特徵，偵測影像或監視錄影是否受到 Deepfake 技術之換臉攻擊，並將建立之 Deepfake Detection 模型部署至自行架設之網站，以提供社會大眾使用。

在進行 Deepfake Detection 的研究前，必須先擁有強大的 Deepfake 技術，而強大的 Deepfake 技術則必須仰賴品質優良的資料集來做訓練。

研究步驟如下：

1. 蒐集網路上蔡英文總統和唐鳳政委之相關影片做為訓練資料。
2. 對蒐集到的訓練資料進行資料預處理，如影音剪接等。
3. 透過 Deepfake 工具與其他新興深度臉部偽造技術合成 Deepfake 影片。
4. 研究與探索世界頂尖研討會中的各種新興 Deepfake Detection 方法，並判別影片是否為 Deepfake 合成。
5. 架設網站並部署 Deepfake Detection 模型。

五、研究成果及效益

如前文所提到，強大的 Deepfake 技術必須基於優良的資料集來做訓練，為了探究資料品質對 Deepfake 的影響，我們將 Deepfake 之成果兩組，第一組為基於低解析度影片（HD）所生成之 Deepfake 影片，如表一所示。第二組為基於高解析度（4K）影片所生成之 Deepfake 影片，如表二所示。

在研究的過程中我們觀察到，由於低解析度的影片能提供給 Deepfake 工具的臉部特徵資訊相對較少，所以低解析度之 Deepfake 影片在細節部分並沒有辦法處理的很細緻，兩人臉部接縫的畫質不佳且並不貼合，有時會出現將側臉貼到正臉的情況；臉部色彩、陰影、光照角度並不自然；蔡英文總統和唐鳳政委都有配戴眼鏡，在眼鏡鏡片與鏡框的銜接處會出現嚴重的歪斜。不僅如此，對低解析度的影片來說，即使每一幀都能進行換臉，但是串接成影片後，整體影片的臉並不自然，動作並不流暢。

與之相對，高解析度影片能提供給 Deepfake 工具的臉部特徵資訊相對較多，所以高解析度之 Deepfake 影片在細節部分處理的更加細緻，使整體動作更加自然與流暢。

以我們自行合成的 Deepfake 影片以及網路上的公開資料集作為基礎，我們建立了 Deepfake Detection 模型，學習影片中的換臉特徵，並將其部署到我們自行架設的網站上，如圖一與圖二(a)(b)所示。

我們所架設的 Demo 網站可以選取蔡英文總統與唐鳳政委的影片進行換臉，並且基於 Deepfake Detection 模型偵測影片是否受到 Deepfake 換臉，最後會在影片左上角呈現結果偵測結果。

表一：低解析度影片之 Deepfake 結果

| 原圖 | 換臉 |
|---|--|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |



表二：高解析度影片之 Deepfake 結果

| 原圖 | 換臉 |
|----|----|
| | |
| | |
| | |
| | |



圖一：架設 Demo 網站

Upload new File (限 .mp4 .wmv)

Choose models

- ☐ 1. LRNet_Celeb
- ☒ 2. DFDC_1_DFDC
- ☒ 3. DFDC_1_Celeb
- ☒ 4. DFDC_1_F2F
- ☐ 5. DFDC_2_DFDC
- ☐ 6. DFDC_2_Celeb
- ☐ 7. DFDC_2_F2F
- ☐ 8. DFDC_3_DFDC
- ☐ 9. DFDC_3_Celeb
- ☐ 10. DFDC_3_F2F
- ☐ 11. Lips_Don't_Lie_DFDC
- ☐ 12. Lips_Don't_Lie_Celeb
- ☐ 13. Lips_Don't_Lie_F2F
- ☐ 14. RFM_DFDC
- ☐ 15. RFM_Celeb
- ☐ 16. RFM_F2F

Choose File 1.mp4

上傳影片

DFDC_1_DFDC: 1.mp4, Prediction label: Real

DFDC_1_Celeb: 1.mp4, Prediction label: Fake

DFDC_1_F2F: 1.mp4, Prediction label: Fake

Final label: Fake

圖二(a)：Deepfake Detection 模型

圖二(b)：Deepfake Detection 結果

六、困難、突破、及未來規劃

如前文所提到，Deepfake 影片的品質取決於我們所蒐集到的素材解析度，低解析度的影片能提供給 Deepfake 工具的臉部特徵資訊相對較少，所以低解析度之 Deepfake 影片在細節部分並沒有辦法處理的很細緻。對低解析度的影片來說，即使每一幀都能進行換臉，但是串接成影片後，整體影片的臉並不自然，動作並不流暢。未來會繼續蒐集更多高解析度影片作為資料集，藉此提升 Deepfake 影片的品質。

目前我們所建立起的 Deepfake Detection 模型雖然可以判別影片是否受到 Deepfake 換臉，然而模型整體的準確度並不高，仍會出現將正常影片誤判為 Deepfake 影片或是相反的情況，Deepfake 和 Deepfake Detection 的研究在學術界正如火如荼的展開，我們將探索更多世界頂尖研討會中全新的 Deepfake 工具以及 Deepfake Detection 模型，藉此提升整體模型的準確度，同時避免現有的 Deepfake Detection 模型無法偵測新形態 Deepfake 影片。