校名：彰化師範大學

人工智慧倫理課程期末實作專題報告

(**斑馬線-行人檢測**)

專案類型：圖片專案

組員名單與貢獻表（務必標註「本人」）

|  |  |  |
| --- | --- | --- |
| 姓名 | 貢獻說明 | 分數(1-5分) |
| XXX | 蒐集訓練資料、訓練模型、簡報報告 | 5 |
| XXX | 蒐集測試資料、測試模型、簡報報告 | 5 |
| XXX | 蒐集訓練資料、倫理分析、簡報報告 | 5 |
| XXX | 蒐集測試資料、訓練模型、 簡報製作、短影音錄製 | 5 |

（本人）

**＃AI使用聲明欄**

o本報告**未**使用任何生成式AI（如ChatGPT, Claude, NotebookLM等）。

**＃學術誠信聲明欄**

o我了解本報告如有抄襲他人著作、未填寫AI使用聲明、有使用AI但未附上對話或截圖、或與校內外他組專題在內容或文字陳述上相似度極高，一律零分計算。

1. **專題主題**

我們的模型旨在希望解決在道路上可能撞傷路人的問題，在現實中，諸如自駕車、輔助駕駛，都有使用到類似的辨識模型。

1. **資料類型與特徵敘述**

本模型使用圖片作為訓練的資料型態，所使用的照片為組員自行拍攝或是從網路上蒐集/截圖而來，故圖片本身有經過人工挑選，確保挑選出的照片畫面的斑馬線、行人清晰可見。

資料的分類依據是判斷斑馬線的附近是否有人存在，可能隱含的偏見是在於資料集的不完整，針對不同的場景(ex:雨天、夜晚)，或是對於不同的族群(ex:輪椅、嬰兒車)，可能會因為資料較少而產生判斷錯誤。

1. **模型設計與資料處理決策**

資料中部份為自己拍攝，部分為網路收集，對於判斷類別的部分則是根據肉眼判斷。

模型設計總共有兩個分類，分別為「馬路上有人」與「馬路上無人」，我們總共訓練了三次，第一次「有人」類別的圖片數量是十八張，「無人」類別五十張，訓練出來模型由於資料量不平衡導致有較多誤判情況，因此決定訓練第二次；第二次，我們將「有人」類別的資料集複製了三分作為新的訓練集，兩個類別各使用五十四張圖片，訓練出了第二版，此時「有人」類別的資料有許多都在相同的路段，畫面十分相似，導致即使是同路段的「無人」照也易被判為「有人」。

由於訓練資料這時只包含白天的狀況，訓練出來的模型在遇到雨天、夜晚的時候就容易判斷錯誤，準確率不佳，在第三次訓練的時候，我們將在訓練集當中重複、類似的圖片剔除，避免讓模型對特定樣貌的路段有所偏心，重新找了新的圖片，加入了更多種不同狀況的圖片（不同時間點、不同天氣），讓模型能接觸到各種不同類型的狀況。我們篩選掉了人物在馬路邊緣行走（非過馬路），以及人物過於模糊的圖片，避免過多的要素影響模型的判斷。

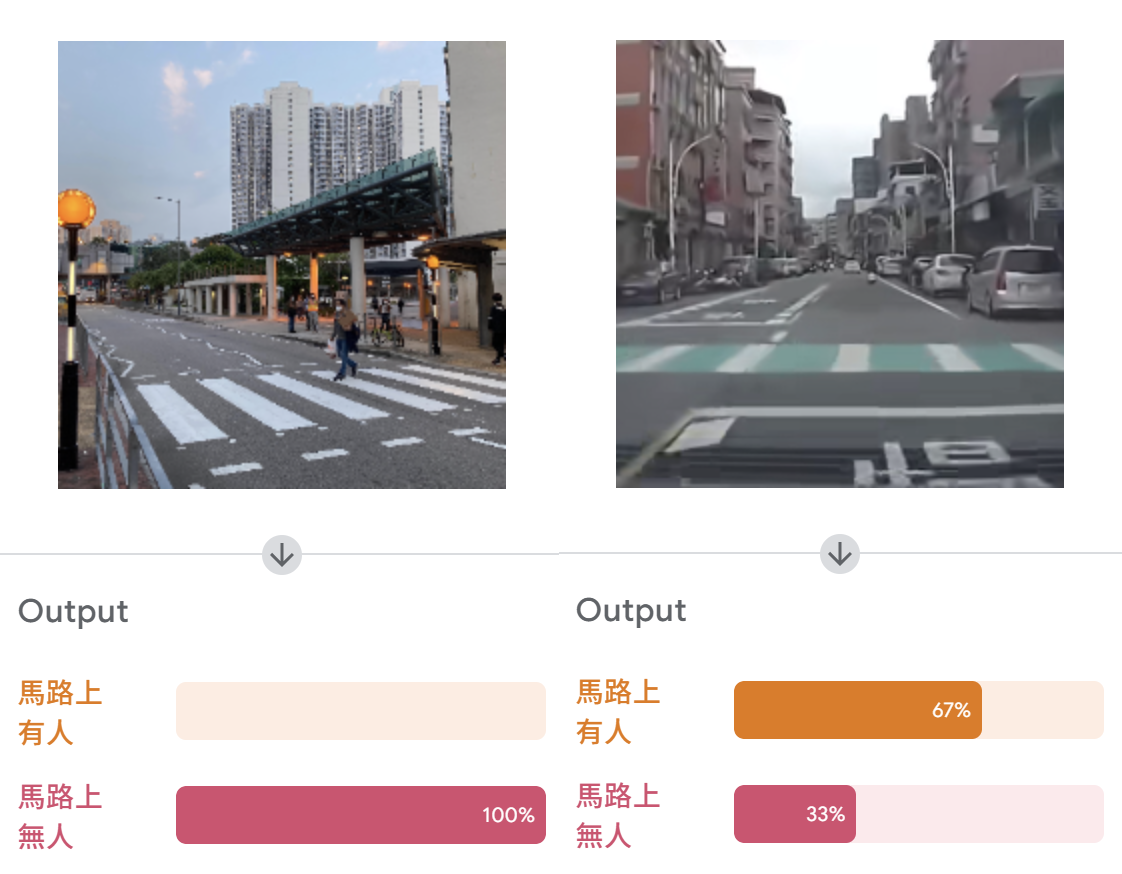
因為Teachable Machine會將圖片自動裁剪，因此在前兩次訓練時，實際上有將類別分到錯誤的地方（以下簡稱裁斷錯誤），我們也將這些錯誤分類圖片事先裁剪，讓訓練的時候能看到正確的圖片樣貌。

這時的測試集涵蓋較多情況的場景，各有五十張圖片，最後訓練出的準確率也因此有所提升。

1. **預測結果表現與誤判觀察**

我們從蒐集的資料分割出有人與無人各三十二張當作測試集，其中預測結果將有人判別為無人有三張，將無人判別為有人的有兩張，其餘的預測結果皆正確，準確率大約落在92%。

基本上誤判情況分為兩種：將無行人判斷成有行人，以及較嚴重地有行人判定成無行人。觀察將有行人判定成無行人的錯誤情況，我們發現當路人的衣著與背景顏色相近時，較容易被判定錯誤（圖一左）；而無行人被判斷有行人的情況，依照觀察模型對同樣圖片的不同畫面裁切的判斷結果，我們得出這張是模型將遠處的機車誤判為行人而導致（圖一右），但測資中其他同樣有機車的圖片都有被正確判斷，因此猜測是偶發性失誤。



圖一、模型偵測結果錯誤示例

1. **模型誤判之倫理分析**
   * 1. **模型誤判可能造成哪些風險與傷害？誰會受影響？**（可對應W9）

* **若此模型應用於真實場景，誤判可能對誰造成損害？**　　當此模型將有人誤判為無人的情況下，可能會使車輛與行人發生碰撞，這樣的行為會對駕駛、乘車人員、行人造成身體、心靈或是生命上的危害。

以下針對模型誤判有人為無人的傷害，從「自由與自主」的原則來分析，透過AI模型來代替駕駛人判斷路上是否有行人，並分析對於駕駛人造成有關思考與做決策的能力的影響。

　若將無人判定為有人，可能會因為汽車根據模型給的預測結果而進行急煞或減速，從而導致乘車人員受到些微的輕傷，或是乘車體驗的不適，亦或是移動時間的延長。

　理論上，有獲取駕照的駕駛人對於看到行徑路線上有行人經過時，應該都會有踩下剎車的下意識舉動，AI應該做為駕駛人精神不集中的情況下的輔助系統，於駕駛人來不及反應的情況下自行啟動剎車作為保險，因此導致駕駛人判斷路上是否有行人的能力下降應該是不太會發生的情況，不過一旦過度信任模型，便會導致駕駛人判斷行人的能力反應速度下降。

　以上的推論會產生另一個問題，「信任」的原則問題，使用者應多大程度上信任AI，雖然使用者本身就具有一定的辨識能力，但不排除有使用者過於信任AI從而導致在駕駛途中完全不確認路線上有無行人，因而導致車輛與行人發生碰撞。

* **有哪些利害關係人？誰應該承擔法律責任或道德責任？**

　　假如此系統用於輔助/自動駕駛上，則受利者為開發自駕車輔助系統的公司高層，和對於自駕車有需求的人，受害者則是部分因為自駕系統而失去工作的司機，和因為自駕系統的判斷失誤而受傷的用路人。

　　使用「責任」原則來分析這次的事故所產生的責任，其中開發團隊內的每一個人、公司中對此專案有下決策或執行的人，和駕駛汽車與車上的乘客均有道德責任，而駕駛人的責任是最重的，因駕駛人過度信任AI模型的判斷，疏忽於自主判斷，連基本的完全義務都未達成。

而法律責任的部分，開發自駕車的公司因模型失誤可能會承擔部分法律後果，但主要是由駕駛人所造成的過失傷害/致死。

除此之外，在因果責任上，是因為模型的開發者的設計疏失和駕駛人的疏忽從而導致意外的發生。

* **這些錯誤是否可預防？錯誤是偶發的還是設計使然？**

是，這些錯誤可以預防，分成兩個層面，一個是設備上的優化：對於模型的優化，使它錯誤機率降低，或是在汽車上增加對於前方障礙物的紅外線感測，皆可預防錯誤。另一層面是人員上的，要求駕駛不得完全信任自駕系統，於駕駛期間仍應注意前方之用路狀況來避免發生意外。

而錯誤的種類同時有偶發與設計使然兩種。如果是相對常見的畫面（ex:白天有成年的行人經過馬路），但出現判斷錯誤，則這屬於偶發性的錯誤，有一定的機率發生；若是出現資料集上相對較少的畫面（ex:光線不佳且畫面模糊，有輪椅經過斑馬線）而產生判斷錯誤，則這屬於設計使然的錯誤，成因是模型的泛化性不夠，對於現實中的稀少族群產生了偏誤，從而導致了模型的判斷錯誤。

* + 1. **為什麼應該降低這些風險或傷害？**（可對應W3/W4/W5）

從效益主義的角度來看，行人辨識的目的是為了提升道路安全和駕駛效率，假如模型運作良好，便能降低車禍發生率，讓社會利益最大化，當模型誤判的時候，可能會造成嚴重的災禍，特別是在「誤判有人為無人」的情況，這樣的代價遠高於它可能節省的時間和資源，所以，從效益主義的角度來看，降低這些風險和傷害，是為了讓社會利益最大化的必要行為。

* + 1. **我們可以如何避免或降低風險？**（可對應W6/W8/W10/W11/W12）
* **在資料蒐集、標註、分類設計上可以如何改進，而讓分類結果更有代表性或公平？**

我們可以拍攝更多不同族群，蒐集更多不常見的情況，但依舊有其他可能出現的情況，拍攝的行人類別也應該豐富越好，舉凡不同年齡，不同膚色，不同外表，讓資料涵蓋的行人族群更廣，降低模型只學會某幾種樣貌的機率，同時手動檢查模型輸出是否出現特定偏好，讓其他人一起測試，從不同視角找出偏誤。

* **若此模型要實際落地應用，有哪些倫理或制度上的建議（例如應遵守哪些倫理原則、法規、審查流程、透明度、問責機制或管理策略）？**
  + 模型可能只會判定特定族群、年齡、穿著，導致對其他人辨識失準。
  + 拍攝到的圖片中可能含有個資，應該要取得他人的同意。
  + 需要避免模型學到個人有關的資料。
  + 要讓開發的過程公開透明，說明資料來源，以及訓練方式。

1. **我們最欣賞的3項外校作品及可學習之處。**
2. 臺灣大學第二組: 臺灣道路交通辨識系統（分為警告∕禁制∕指示∕輔助）

這組的可學習之處在於，他們有想到「讓使用者作為資料提供者」的思路，讓這個用於辨識交通號誌的模型能隨著時間推移，看到更多種情況，藉此提升準確率

1. 高雄醫學大學第三組（分為背景噪音∕AI合成人聲∕真實人聲）

這組的影片是先拋出一個現實存在的問題，讓聽眾得知這個模型有存在的必要，再去講述這個模型的訓練細節，他們的資料集也不僅限於中文，還包含多個podcast的內容，讓準確度變得更高一些。

1. 臺中科技大學第二組: AI協助燙傷分級判讀（一級∕二級∕三級燙傷∕其他）

他們有提供使用者信心指標的判斷結果，並明確指出該模型，應僅作為輔助醫療診斷的工具；還預想到使用者不了解醫療相關知識的情況，並提供分級解釋。