

適用於擴增實境應用之擬真打光系統

Photo-Realistic Lighting System for AR Applications

專題組員：吳昱萱、陳佰賢、陳玥華、余忻璋

專題編號:PRJ-NTPUCSIE-111-007

執行期間：2022年6月至2023年6月

1. 摘要

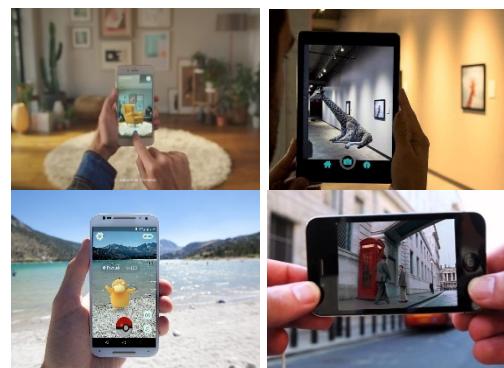
隨著軟體與硬體的進步，能夠結合虛擬物體與真實世界環境的擴增實境技術（Augmented Reality，以下簡稱 AR）在日常生活有愈來愈多的應用。為了增加畫面的真實性，AR 應用在渲染虛擬物體時需使用與真實世界的光源設定，才不會因畫面的不一致而破壞沉浸感。本次專題擬設計一套適用於 AR 應用的擬真打光系統，藉由設計光源偵測與分類神經網路、結合前人的研究、整合擴增實境套件、以及實作提升數位視覺效果的電腦圖學渲染器，我們的系統能更準確地重建真實世界的光源以提升虛擬物體的渲染品質，提升擴增實境應用中畫面的真實性與沉浸感。

2. 簡介

(1) 研究動機與問題

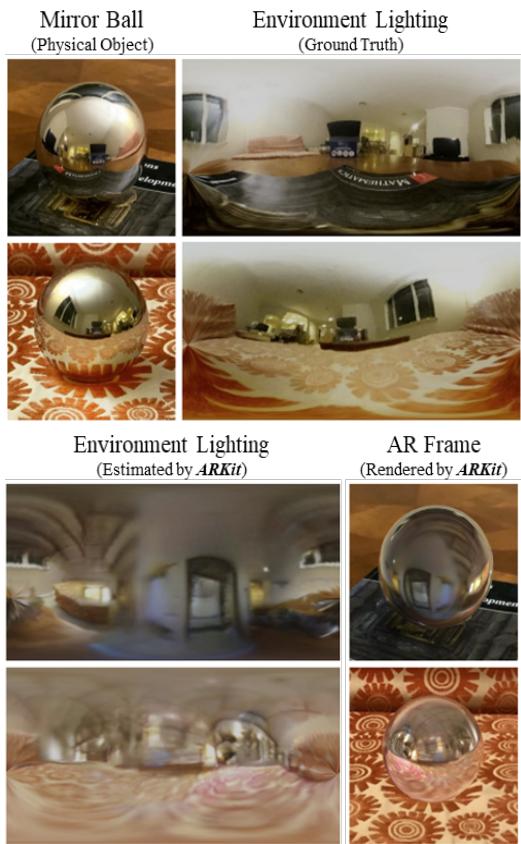
在2021年底 Facebook 更名為 Meta 並宣布其元宇宙的計畫後，作為元宇宙入口的延展實境技術（Extended Reality，XR）不但成為媒體聚光燈的焦點，也成為科技業及學術界致力研究的對象。延展實境為虛擬實境（Virtual Reality，VR）、擴增實境（AR）、混合實境（Mixed Reality，MR）三者的總稱。相較於虛擬實境技術企圖以程式創造一個完全的虛擬世界，在真實世界的畫面中融合虛擬物件以及數位資訊的擴增實境以及混合實境技術，能更大程度的影響我們

的生活。[圖一](#)中展示了許多擴增實境應用，包括傢俱挑選與擺設、博物館的互動式導覽、手機遊戲、以及藝術創作。



圖一：擴增實境技術在日常生活中的各種應用。從左上至右下為傢俱擺設、互動式導覽、手機遊戲、及藝術創作。圖片取自網路。

由於在 AR 與 MR 的畫面中會同時出現真實環境與虛擬的物體，如何讓虛擬物體看起來像處在真實環境中，就成為一個非常重要的研究課題。目前大多數的 AR 應用能藉由偵測場景中的二維條碼（例如 QR Code）或是平面區域將虛擬物體穩定的放置在正確的位置，但在真實光源的估計以及虛擬物體的繪製上，則仍有許多進步空間，[圖二](#)中展示了 ARKit 套件[1]的光照結果，可以看出其估計出的環境光源與真實世界有很大的差異，因此繪製虛擬物體時會產生與真實環境不一致的光影及反射。鑑於此原因，本計畫擬開發一套適用於 AR 應用的擬真打光系統，以提升擴增實境應用中虛擬物體的逼真程度、增進使用時的沉浸感。

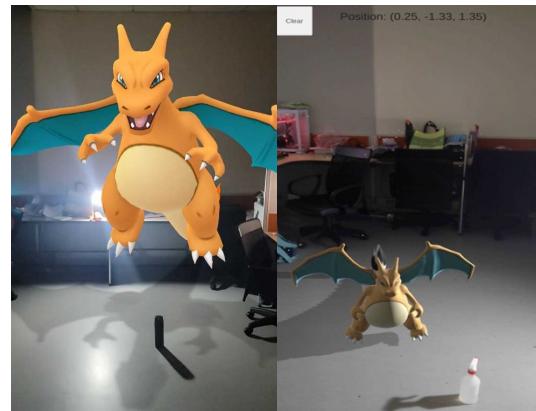


圖二：ARKit 套件[1]的光照結果。上圖左欄為兩個以金屬球反射捕捉真實世界光源的例子，上圖右欄為使用捕捉的結果生成的環境光源，可視為標準答案。下圖兩欄分別為使用 ARKit 套件在同樣環境下估計的環境光源，以及繪製虛擬物體的結果。可以看出由學習演算法猜測的環境光源與真實世界的光源有很大的差距。圖片修改自論文[2]。

(2) 相關研究

過去有許多研究嘗試利用畫面中真實物體的顏色重建真實世界的環境光源，例如金屬球[3]、眼球[4-5]、人臉[6-8]、或是已知幾何形狀的物體[9]。由於上述的方法需要場景中具有特定的物體，對於實際的擴增實境應用有諸多限制。近年來由於深度學習的快速進步，許多研究以學習演算法重建真實世界光源。Gardner 提出的卷積神經網路（convolutional neural network），能以單張小視野影像還原完整環境光源[10]。LeGendre 等人藉由訓練時期在場景中擺設三種不同材質的圓球，觀察不同反射率的圓球在相同光源下的光照結果，訓練重建光源的類神經網路[11]。Garon[12] 以及 Marques[13] 等人一樣使用卷積神經網

路由一張場景照片回推場景光源，其



圖三：本計畫之主要成果。藉由與 Pokémon Go 的打光系統作比較（左圖），可以看出本系統（右圖）能大幅提升虛擬物體的渲染效果，及產生出正確的陰影。

類神經網路能直接生成環境光源的球諧函數，因此除了較容易訓練之外，也能夠直接使用於實時繪圖的應用中。

值得注意的是 Gardner 的論文[10] 中也有使用神經網路偵測 LDR 全景圖中是否出現聚光燈、檯燈、窗戶、或反射等類型的能量，然而其目的為提升 HDR 全景圖的生成品質，並沒有將這些光源根據特性做不同的渲染處理。

(3) 目標與預期效益

目前對於三維世界的光源描述大多使用 360 度全景圖（360 Panorama）格式，用以描述球面坐標系下各個方向入射至場景的能量。本專題亦使用單張 360 度全景圖作為輸入，隨著行動裝置的進步，一般使用者已能很輕易地透過手機拍攝 360 度全景圖。同時，若場景光線變動不大，此拍攝對於同一場景僅需進行一次，因此並不會造成使用者太大的負擔。

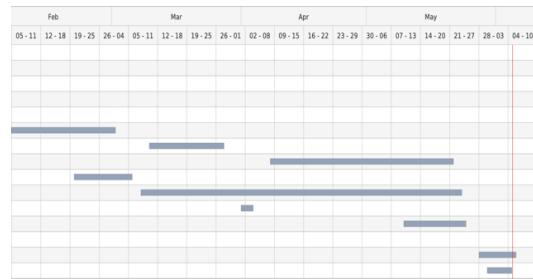
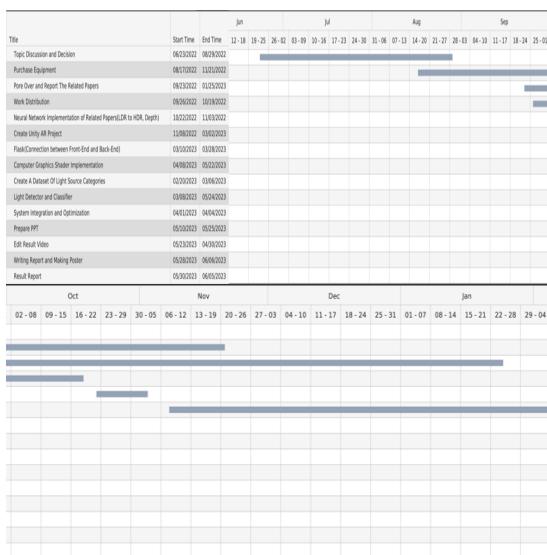
本研究設計的系統能由單張 360 度全景圖產生適用於 AR 應用的光源資料，其主要包含兩部分：第一，由於直接使用全景圖打光的計算成本太高，本系統採用與大多數 AR 系統相同的方法，將 360 度全景圖轉換至球諧函數基底（Spherical Harmonics，SH）以達成即時繪圖的目標。第二，由於 SH 僅能產生低頻率的光照效果，

且無法生成陰影，本計畫的最大主要貢獻為開發出一個深度學習網路，能從單張360度全景圖中偵測與分類主要光源（Key Lights），這些主要光源能提供高頻的打光效果以及陰影。

最後，我們將生成的光源資料整合至遊戲引擎與 AR 套件，實際測試其畫面效果的提升。本專題還利用電腦圖學課程學到的知識，開發出繪製陰影效果的 GPU 著色器 (Shaders)，更進一步提升陰影的繪製。圖三為 Pok  emon Go 的打光系統與我們系統的比較圖。從左邊的圖片可以發現 Pok  emon Go 中的影子與水壺實際的陰影相去甚遠，而右圖中我們的系統則能產生與現實中顏色、方向更為一致的影子，虛擬物體的著色亦更為真實。

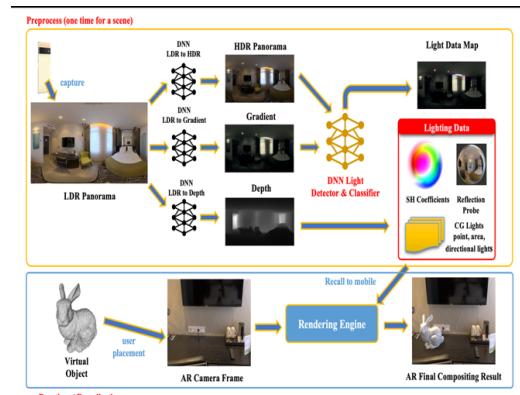
3. 專題進行方式

受限於手機端的計算效能，在手機上產生光照資料將會花費太多時間，降低使用者的體驗。為了解決此問題，我們選擇以電腦、Android 手機作為實作的呈現平台。後端以 Python 開發深度學習網路及生成虛擬光源，前端使用 Unity 遊戲引擎與 AR 套件將計算得到的光照資料整合至 AR 應用中，中間的連接則是利用 flask 框架來實現。我們以兩人為一組分別處理前端和後端，前端的部分是由陳佰賢、吳昱萱負責，而後端的部分則為余忻璋、陳玥華負責。**圖四**為本計畫的時程規劃。



圖四：本專題之時程規劃，皆有按時完成。

圖五為本系統之架構圖，可以分為前處理（Preprocess）及擴增實境的即時應用（Run-time AR application）兩部分。圖中黃色區塊為前處理的部分，其目的為重建使用者當下的環境光源。如先前所提到的，我們假設使用者以個人手機或平板取得環境的全景圖，作為本系統之輸入。由於真實世界中的光照資料為高動態範圍（High Dynamic Range，HDR），但目



圖五：本系統架構圖，可以分為前處理（黃色邊界區塊）及即時應用（藍色邊界區塊）兩部分。在前處理部分我們使用三組類神經網路得到環境的光源，每個場景僅需執行一次。前處理得到的光源可應用於AR即時應用中以提升虛擬物體的繪製品質。

前大多行動裝置僅能取得低動態範圍（Low Dynamic Range，LDR）的影像，我們首先使用[\[14\]](#)的類神經網路由 LDR 全景圖生成 HDR 資料。另外，為了模擬區域光源造成的不同光影效果，我們還整合了 Alhashim 與 Wonka 的方法[\[15\]](#)估計全景圖的深度，以還原虛擬光源的三維位置。

接著我們將 LDR 的與 HDR 全景圖同時輸入本計畫開發之光源偵測與分類神經網路，以分析場景中光源的類型與位置。我們的系統原本直接使

用 Yolo v7 [16]來偵測光源，但在訓練的過程中發現，由於世界上的光源種類非常多，且具有不同的外型與顏色（有些甚至不具有明顯的外型），再加上 Yolo 並不是針對光源偵測設計的網路，因此即使在以我們標註的資料重新訓練後，準確度依然不佳。為了解決此問題，我們先由 HDR 全景圖中計算出一張梯度圖，將此一同放入 Yolo 中進行訓練。當梯度大時表示光源集中，應設定為點光源，若梯度相對小則設定為區域光源。經過實驗證實加入梯度圖後能大幅提升模型偵測光源以及判斷種類的準確率，由 0.48 提升至 0.67。圖六展示了本系統點光源與區域光源的梯度圖。

光源偵測與分類器的輸出為一張光源遮罩 (Light Data Map)，其描述全景圖中所有直接光源的種類與位置



圖六：點光源（左）與區域光源（右）的梯度圖。

(方向)，後續系統則依照種類設定不同的電腦圖學光源，目前我們區分了點光源及區域光源，不屬於直接光源的部分則會被歸類為間接光源，以球諧函數 (SH) 及反射探針 (Reflection Probe) 表示其漫反射及似鏡面反射的能量。值得一提的是由於目前缺乏能在行動端即時繪製任意形狀區域光源的繪圖演算法，我們實作由 Debevec 提出的 Median Cut 演算法[17]，將區域光源轉換成若干個點光源進行模擬，在繪圖品質及計算效能上取得一個很好的平衡。

在完成場景光源的重建後，我們將生成的光源資料（包含 SH 系數、反射探針、以及重要直接光源的位置及能量）整合於現有的 AR 套件，實際比較其帶來之畫面品質的提升。我

們還實作數個電腦圖學著色器，進一步提升陰影的繪製效果。

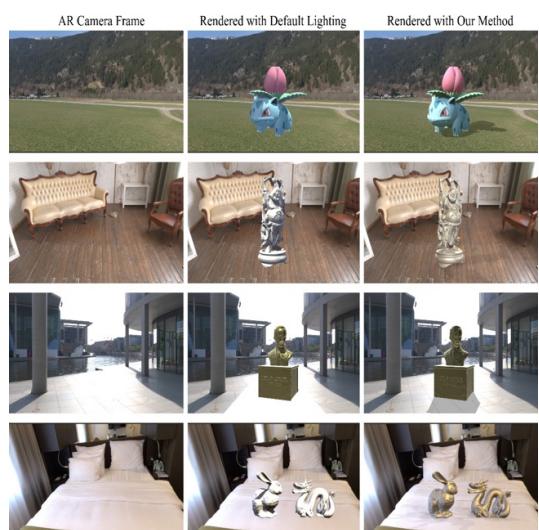
除了上述遇到的手機計算效能問題，在專題進行期間我們還遇到了幾個困難：首先，因為目前並不存在具有光源類別的資料集，所以我們需要自己花費時間來手動標記點光源和區域光源。第二，Unity 中的座標轉換和熟悉的笛卡兒座標並不一樣，所以我們花了不少時間對現實中的座標與方向做數學對應，使其能在 Unity 中正確的呈現。

4. 主要成果與評估

圖七展示了本計畫開發之光源偵測與分類器之測試結果。左圖為輸入的 LDR 全景圖，中圖為我們使用 [3] 的方法得到的 HDR 全景圖，右圖則為光源偵測之結果。由結果可見大部分的主要光源皆能成功的偵測與分類。



圖七：目前 HDR 全景圖和 Light Data Map 的生成結果。左為輸入的 LDR 測試全景圖，中間為使用 [14] 的方法得到的 HDR 全景圖，右為我們系統產生的 Light Data Map。



圖八：本系統預期成果比較，此比較使用網

路上現有的 HDR 全景圖，以電腦圖學的方式在遊戲引擎中的模擬得到。比較中間與右欄可以看見我們的系統應能大幅增進虛擬物體的真實程度。

圖八為將偵測到的光源用於繪製虛擬物體的成果，為了方便比較，此圖使用的是我們使用 Unity 開發的桌機版本 AR 模擬器，其使用 360 全景圖作為背景，並以一台虛擬相機模擬 AR 畫面的輸出（左）。中圖為使用 Unity 預設光源繪製虛擬物體的結果，由於沒有套用跟真實環境一致的光源設定，虛擬物體看起來並不真實。右圖為使用本系統重建之光源繪製虛擬物體的結果，虛擬物體的顏色與背景明顯較為融合，再加上我們開發的陰影著色器後，得到了非常逼真的效果。

圖三展示了本系統最終整合至 AR 應用的結果。由這些例子可以看出藉由捕捉與重建真實世界的光源，虛擬物體的逼真度能顯著的提升，大幅增進了使用者進行 AR 體驗時的沉浸感。其餘結果請參見影片。

最後，本專題之工作項目及貢獻條列如下：第一，創新部分：我們結合傳統方法以及類神經網路，設計出一個能夠偵測及判斷全景圖中各種不同類型的光源位置與形狀的演算法。第二，演算法整合部分：我們實作兩篇前人的研究，首先以 Eilertsen 等人提出的類神經網路[14]將手機拍攝得到的 LDR 全景圖轉換成 HDR 的版本；並以 Alhashim 與 Wonka 的方法[15]估計全景圖的深度，用以判斷光源的遠近。最後是系統整合部分：我們在 Unity AR 開發套件[18]中使用前處理估計出的光源資料，以驗證系統之成效。同時，我們亦在 Unity 遊戲引擎中實作多個電腦圖學渲染著色器（shader），進一步的提升畫面效果。

至於本專題的預期與實際成效有兩個主要的差異：第一，由於資料集數量的不足，因此光源偵測與分類器無發分辨出平行光與聚光燈，僅能辨別點光源以及區域光源。第二，我們

的深度神經網路目前較適用於室內光源的偵測，室外場景的結果較不如預期，因此我們希望未來可能之擴展方向以克服這兩項差距為主。

5. 結語與展望

隨著科技的普及相信未來在 AR 上的應用也會越來越廣，我們的系統透過將光源分類來做光影變化的方法將使得虛擬物件與現實世界中的互動變得更加擬真。

在研究中我們提出了兩個嶄新的想法。第一，我們使用 MedianCut 演算法將區域光源以數個點光源來模擬，使得我們在效能與視覺效果上達成了完美的平衡，其不僅解決了目前即時區域光繪圖演算法無法處理複雜光源外型的問題，也能使運算大量降低，讓我們的系統能在各個行動端中做使用，大幅增加了普及性。第二，我們結合梯度和原圖放進 Yolo 神經網路中進行訓練，大幅提升判斷與分類的準確度。在未來若是我們將光源的分類劃分的更加詳細，抑或是在即時的深度預測上有突破性的發展，相信我們可以創造出一個真假難分的 AR 世界。

6. 銘謝

本次專題中，教授的指導內容可以分為以下幾點：

(1) 論文的研讀收集和報告

在定義題目初期，指導教授挑選了許多不同面向的相關論文，讓我們定期報告，以便對這個題目有更深刻的理解。經過大量的閱讀以及資料過濾後，最後確立了明確的方向與主題。

(2) 類神經網路的設計

在設計光源偵測及分類器時，教授與我們討論類神經網路的架構以及設計邏輯，並在遇到問題時分析可能的緣由。

(3) 區域光源的取樣

擴增實境應用通常為即時互動，

因此需要維持一定的幀數 (frame rate)，而區域光源的渲染會花費太多的時間。指導教授的建議可以採用電腦圖學中對大面積光源取樣的演算法，在不損失太多畫面品質的情況下增進繪圖效率。

(4) Unity 的教學

儘管 Unity 的功能十分強大，但因為過於龐大的函式庫，使得初步嘗試的我們遇到許多障礙跟瓶頸，經過指導教授充分解釋 Unity 的系統架構及使用方法後，我們才能很快的進入狀況。

(5) 計畫的校稿以及成果的撰寫

教授除了指導我們學術報告的撰寫格式以及應注意的事項之外，也提點我們撰寫的邏輯，讓整篇報告更通順容易閱讀。

7. 參考文獻

- [1] Apple Inc. 2022. Introducing ARKit 6. <https://developer.apple.com/augmented-reality/arkit/>
- [2] Y. Zhao, C. Ma, H. Huang, and T. Guo, “LITAR: Visually coherent lighting for mobile augmented reality,” in Proc. UbiComp 2022.
- [3] P. Debevec, “Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography,” ACM Trans. Graph., 1998, p. 189–198.
- [4] K. Nishino and S. K. Nayar, “Eyes for re-lighting,” ACM Trans. Graph., vol. 23, no. 3, aug 2004, p. 704–711.
- [5] H. Wang, S. Lin, X. Ye, and W. Gu, “Separating corneal reflections for illumination estimation,” Neurocomput., vol. 71, no. 10–12, jun 2008, p. 1788–1797.
- [6] M. H. Conde, D. Shahlaei, V. Blanz, and O. Loffeld, “Efficient and robust inverse lighting of a single face image using compressive sensing,” in IEEE ICCV Workshop, 2015, p. 226–234.
- [7] D. Shahlaei, M. Piotraschke, and V. Blanz, “Lighting design for portraits with a virtual light stage,” in IEEE ICIP, 2016, p. 1579–1583.
- [8] M. Zollhöfer, J. Thies, P. Garrido, D. Bradley, T. Beeler, P. Perez, M. Stamminger, M. Nießner, and C. Theobalt, “State of the art on monocular 3d face reconstruction, tracking, and applications,” Computer Graphics Forum, vol. 37, no. 2, 2018, p. 523–550.
- [9] D. Mandl, K. M. Yi, P. Mohr, P. M. Roth, P. Fua, V. Lepetit, D. Schmalstieg, and D. Kalkofen, “Learning lightprobes for mixed reality illumination,” in Proc. IEEE ISMAR, 2017, p. 82–89.
- [10] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde, “Learning to predict indoor illumination from a single image,” ACM Trans. Graph., vol. 36, no. 6, nov 2017.
- [11] C. LeGendre, W.-C. Ma, G. Fyffe, J. Flynn, L. Charbonnel, J. Busch, and P. E. Debevec, “DeepLight: Learning illumination for unconstrained mobile mixed reality,” in Proc. IEEE/CVF CVPR, 2019.
- [12] M. Garon, K. Sunkavalli, S. Hadap, N. Carr, and J.-F. Lalonde, “Fast spatially-varying indoor lighting estimation,” in Proc. IEEE/CVF CVPR, June 2019.
- [13] B. A. D. Marques, E. W. Gonzalez Clua, A. A. Montenegro, and C. Nader Vasconcelos, “Spatially and color consistent environment lighting estimation using deep neural networks for mixed reality,” Computers & Graphics, 2021.
- [14] G. Eilertsen, J. Kronander, G. Denes, R.-K. Mantiuk, J. Unger, “HDR image reconstruction from a single exposure using deep CNNs,” ACM Trans. Graph., vol. 36, no. 6, 2017.
- [15] I. Alhashim and P. Wonka, “High quality monocular depth estimation via transfer learning,” arXiv 2018.
- [16] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”.
<https://arxiv.org/abs/2207.02696>
- [17] Paul Debevec, “A median cut algorithm for light probe sampling,” SIGGRAPH 2005 poster.
- [18] Unity Augmented Reality ToolKit.
<https://unity.com/unity/features/ar>