計算機視覺期末報告

3D 領域應用、醫療、路徑規劃

干皓丞 2101212850

CVPR 2021 Best Paper

在此選出 CVPR 2021 的 8 篇文獻進行精讀，同時嘗試進行復現。

Mihai Dusmanu et al.	個人隱私保護
Wei Ji et al.	醫療
Shitong Luo et al.	3D 形狀建模
Jennifer J. Sun et al.	數據行為
Kehong Gong et al.	3D 人體姿勢
Shunsuke Saito et al.	3D 掃描
Lea Müller et al.	3D 人體姿勢
Abhishek Badki et al.	路徑規劃
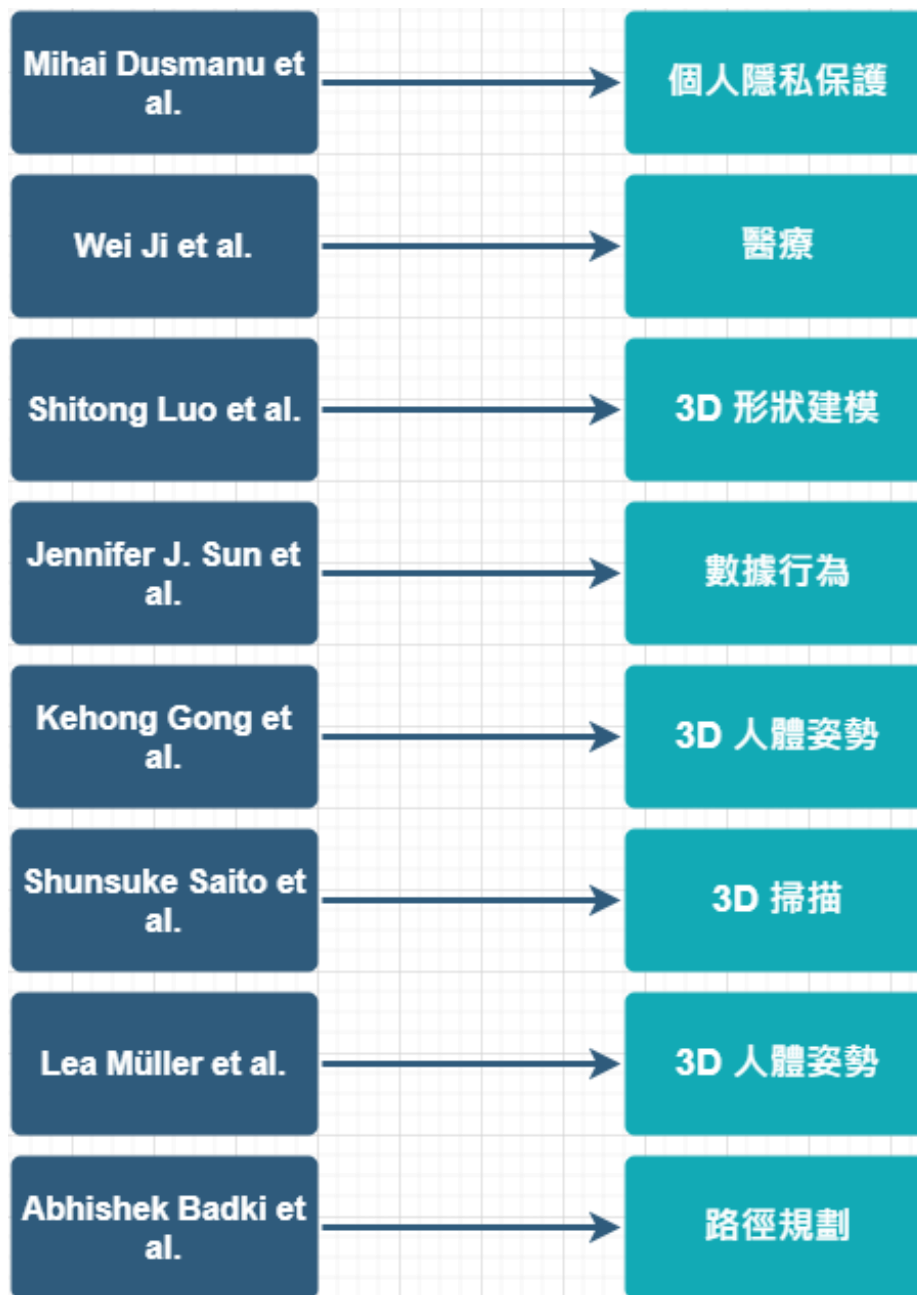
CVPR 2021 Best Paper

許多計算機視覺系統需要將使用者的圖像特徵上傳到雲端進行處理和儲存，而該研究發現可以利用這些特徵來恢復有關場景或主題的敏感資訊，比如通過重建原始圖像的外觀。而研究者們為了解決這個隱私問題，提出了一種新的隱私保護特徵表示。
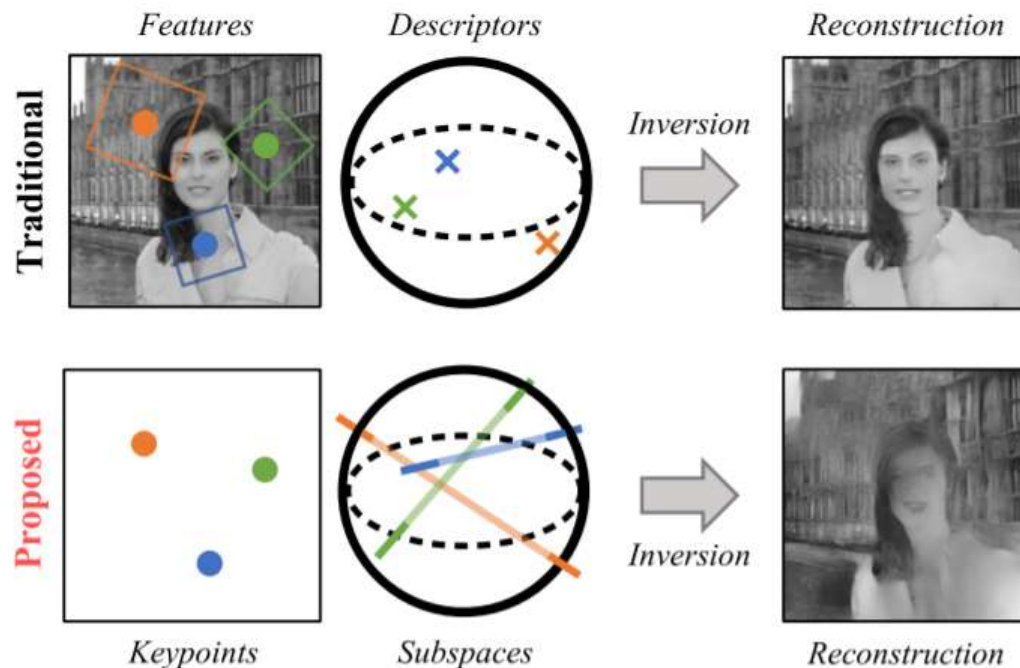

Figure 1: **Privacy-Preserving Image Features.** Inversion of traditional local image features is a privacy concern in many applications. Our proposed approach obfuscates the appearance of the original image by lifting the descriptors to affine subspaces. Distance between the privacy-preserving subspaces enables efficient matching of features. The same concept can be applied to other domains such as face features for biometric authentication. Image credit: *laylamoran4battersea* (Layla Moran).

CVPR 2021 Best Paper

在醫學圖像分析領域中，通常會收集多個註釋，而每一個註釋都來自不同的臨床專家或評估者，以期望減少可能的診斷錯誤。而從計算機視覺的角度來看，採用通過多數票或來自首選評估者的簡單註釋來獲得的真實標籤已成為一種常見做法。
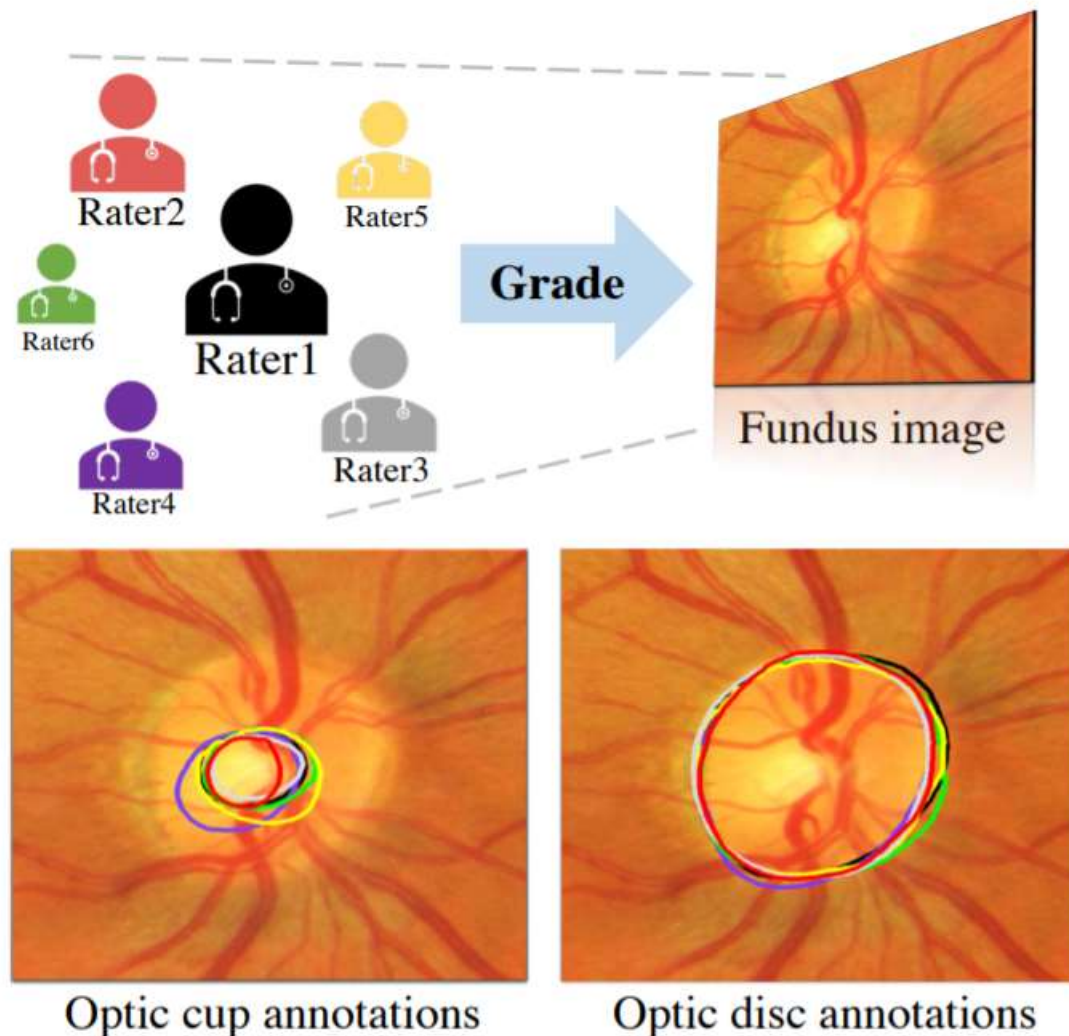

Figure 1. **Top:** an exemplar medical image grading scenario conducted by multiple raters with different expertise levels. **Bottom:** visualization of optic cup and disc annotations of the above raters.

CVPR 2021 Best Paper

研究者提出了一個用於點雲生成的概率模型，該模型是各種 3D 視覺任務的基礎，比如形狀完成、上採樣、合成和數據增強。而受非平衡熱力學中擴散過程的啟發，我們將點雲中的點視為與熱浴接觸的熱力學系統中的粒子，它們從原始分佈擴散到噪聲分佈。因此，點雲生成相當於學習將噪聲分佈轉換為所需形狀分佈的反向擴散過程。
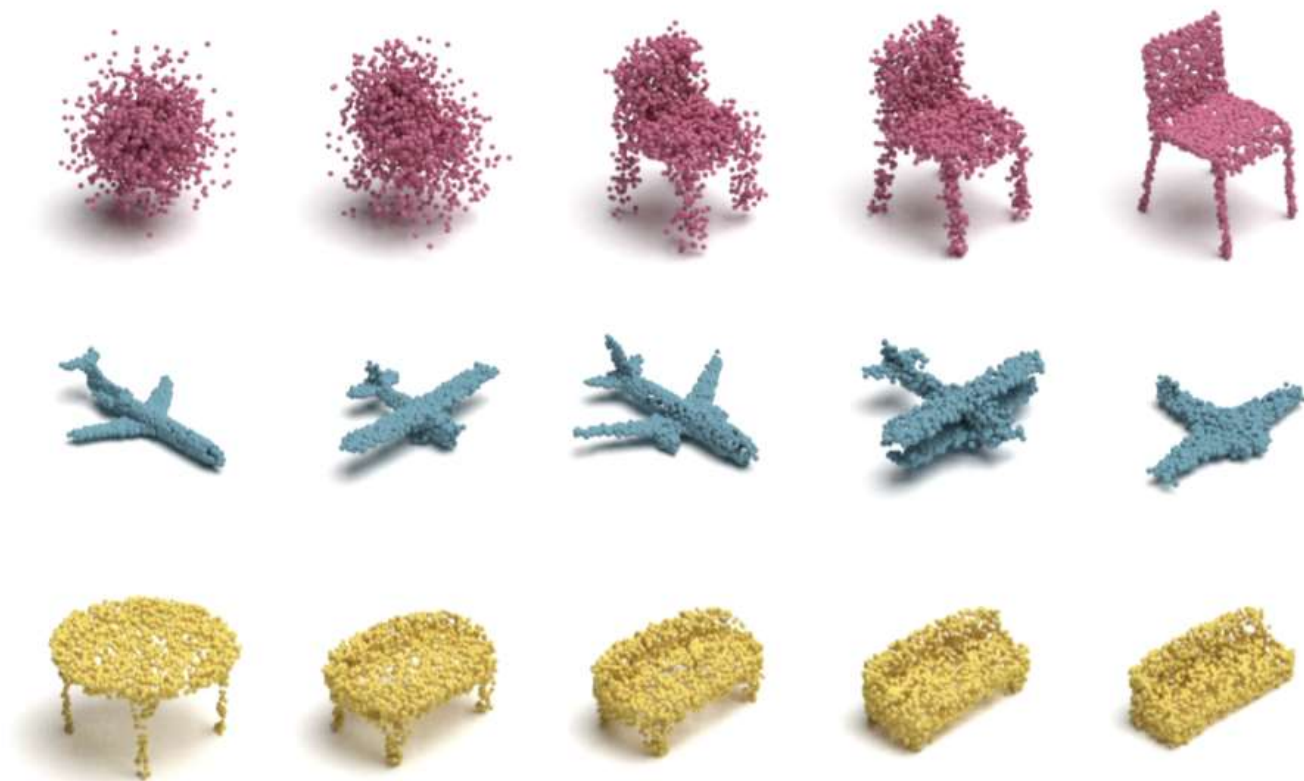

Figure 1. **Top**: The diffusion process that converts noise to some shape (left to right). **Middle**: Generated point clouds from the proposed model. **Bottom**: Latent space interpolation between the two point clouds at both ends.

CVPR 2021 Best Paper

專業領域知識通常是準確註釋訓練集以進行深入分析所必需的，但從領域專家那裡獲取可能既繁瑣又耗時，該問題在自動化行為分析中尤為突出，其中從影片跟踪數據中檢測到 Agent 移動或者是感興趣的動作。為了減少註釋工作，研究者提出了 TREBA，一種基於多任務自監督學習學習註釋樣本有效軌跡嵌入以進行行為分析的方法。
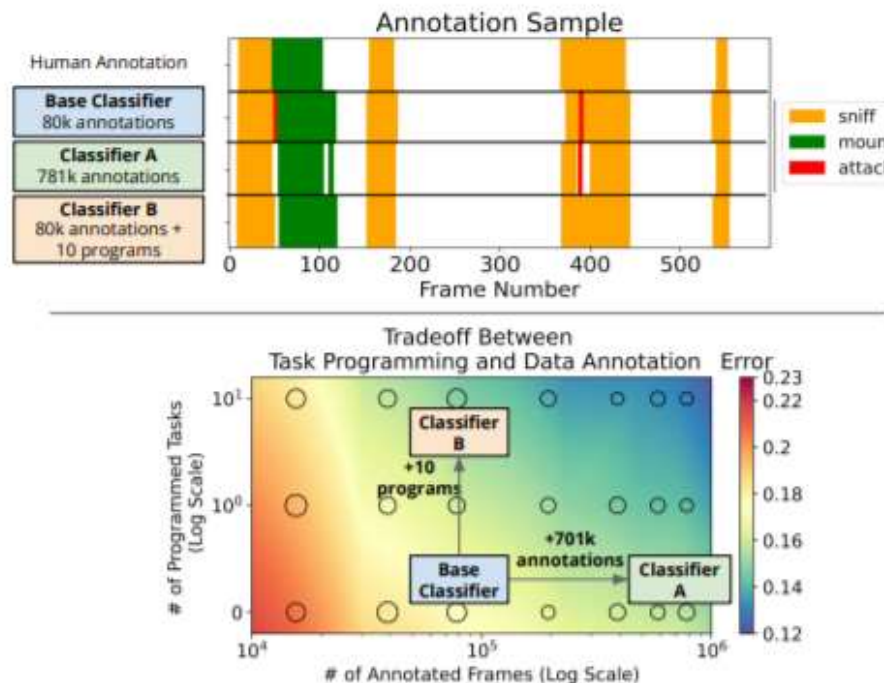

Figure 1. **Overview of our approach.** *Part 1:* A typical behavior study starts with extraction of tracking data from videos. We show 7 keypoints for each mouse, and draw the trajectory of the nose keypoint. *Part 2:* Domain experts can either do data annotation (Classifier A) or task programming (Classifier B) to reduce classifier error. The middle panel shows annotated frames at 30Hz. Colors in the bottom plot represent interpolated performance based on classifier error at the circular markers (full results in Section 4.3). The size of the marker represents the error variance.

PoseAug：用於 3D 人體姿勢估計的可微姿勢增強框架

CVPR 2021 Best Paper

現有的 3D 人體姿態估計器對新數據集的泛化性能較差，這主要是由於訓練數據中 2D-3D 姿態對的多樣性有限。而研究者為了解決這個問題，提出了 PoseAug，這是一種新的自動增強框架，它可以學習將可用的訓練姿勢增加到更大的多樣性，從而提高經過訓練的 2D 到 3D 姿勢估計器的泛化。
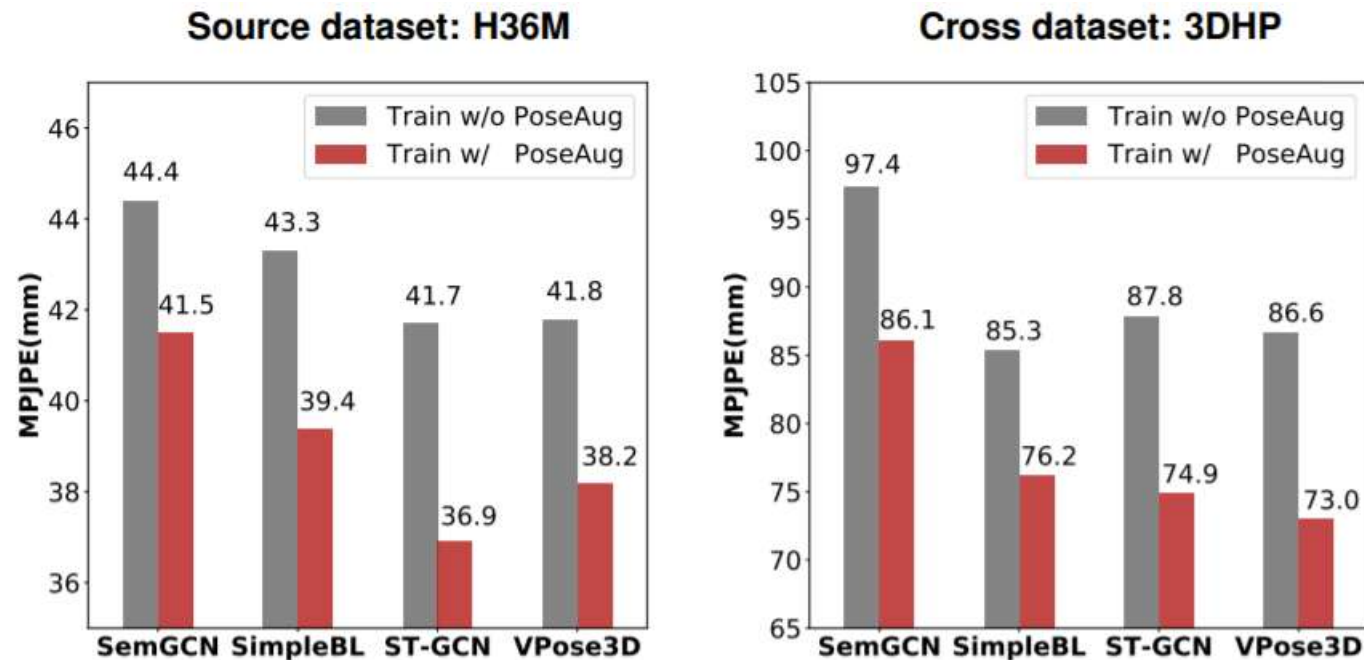

Figure 1: Estimation error (in MPJPE) on H36M (intra-dataset evaluation) and 3DHP (cross-dataset evaluation) of four well established models [52, 26, 33, 3] trained with and without PoseAug. PoseAug significantly improves their performance for both the intra- and cross-dataset settings.

CVPR 2021 Best Paper

研究展示了名為 SCANimate 的一個端到端的可訓練框架，它可對一個穿著衣服的人進行原始 3D 掃描並將它們變成一個可動畫的化身，這些化身由姿勢參數驅動，並擁有可自然移動和變形的逼真服裝。SCANimate 不依賴於自定義網格模板 (a customized mesh template) 或表面網格註冊 (surface mesh registration)。
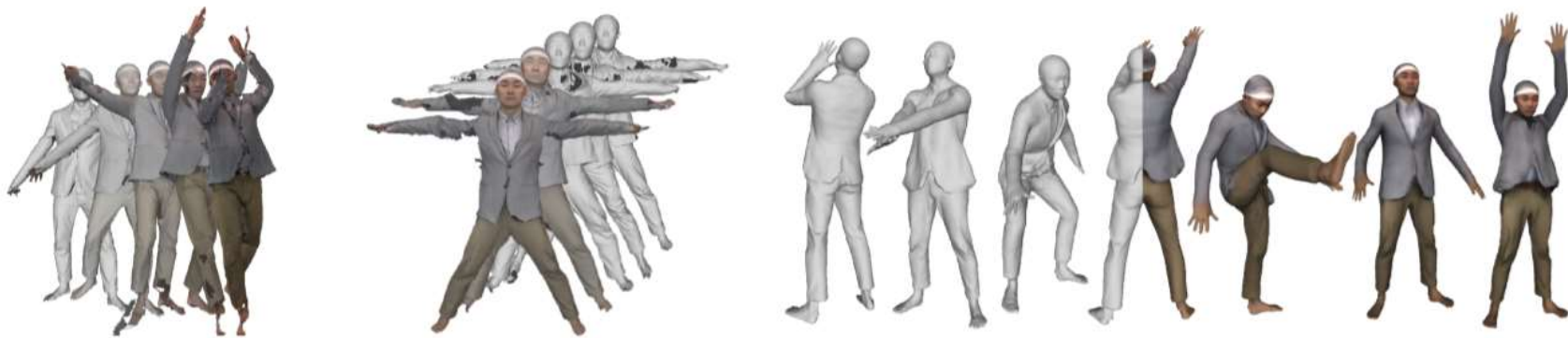

Figure 1: **SCANimate.** Given a set of raw scans with multiple poses containing self-intersections, holes, and noise (left), SCANimate automatically aligns all scans to a canonical pose (middle) and learns a Scanimat, a fully animatable avatar that produces pose-dependent deformations and texture without garment-specific templates or mesh registration (right).

CVPR 2021 Best Paper

人們每小時摸臉 23 次，他們交叉雙臂和雙腿，把手放在臀部等，雖然許多人的圖像包含某種形式的自我接觸，但當前的 3D 人體姿勢和形狀 (HPS) 回歸方法通常無法估計這種接觸，而為了解決這個問題，研究者開發了新的資料集和方法，通過自我接觸顯著改善人體姿勢估計。
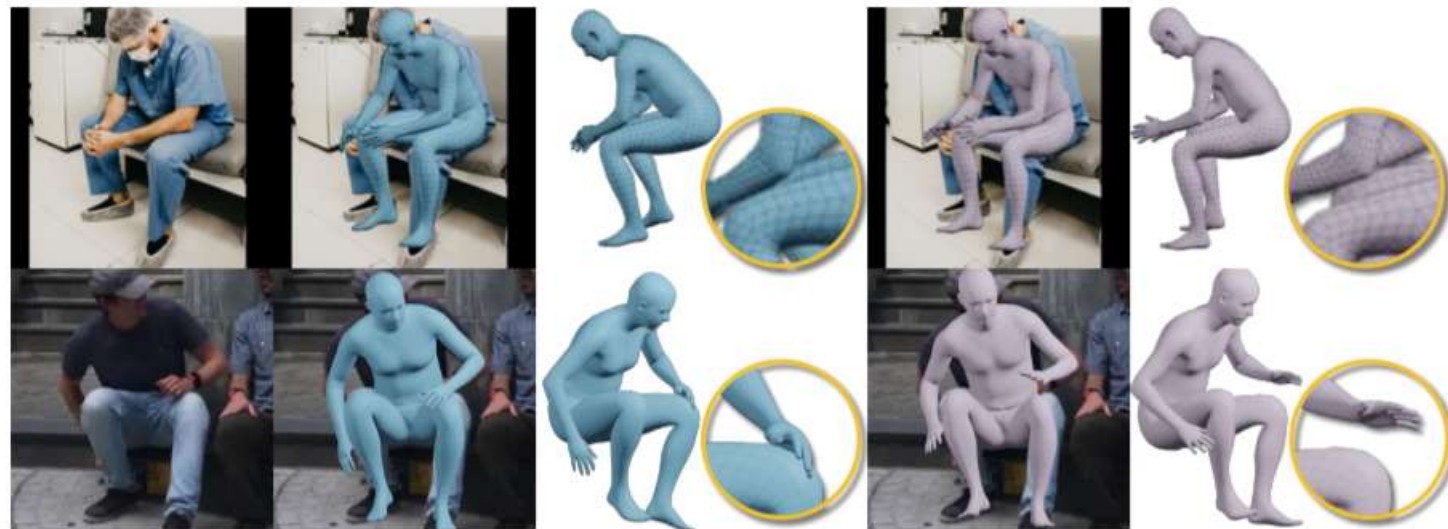

Figure 1. The first column shows images containing self-contact. In blue (left), results of TUCH, compared to SPIN results in violet (right). When rendered from the camera view, the estimated pose may look fine (column two vs. four). However, when rotated, it is clear that training TUCH with self-contact information improves 3D pose estimation (column three vs. five).

CVPR 2021 Best Paper

接觸時間 (TTC)，即物體與觀察者平面碰撞的時間，是路徑規劃的強大工具：它可能比場景中物體的深度、速度和加速度提供更多信息—即使對於人類也是如此。

對於這種情況，該研究的方法在 6.4 毫秒內提供了時間地理圍欄 (temporal geofence)，此方法比現有方法快了 25 倍。當計算預算允許時，該研究的方法還可以通過任意精細的量化，包含連續值來估計每一個像素的 TTC。此方法是第一個以足夠高的幀速率提供 TTC 資訊 (binary or coarsely quantized) 以供實際使用的方法。
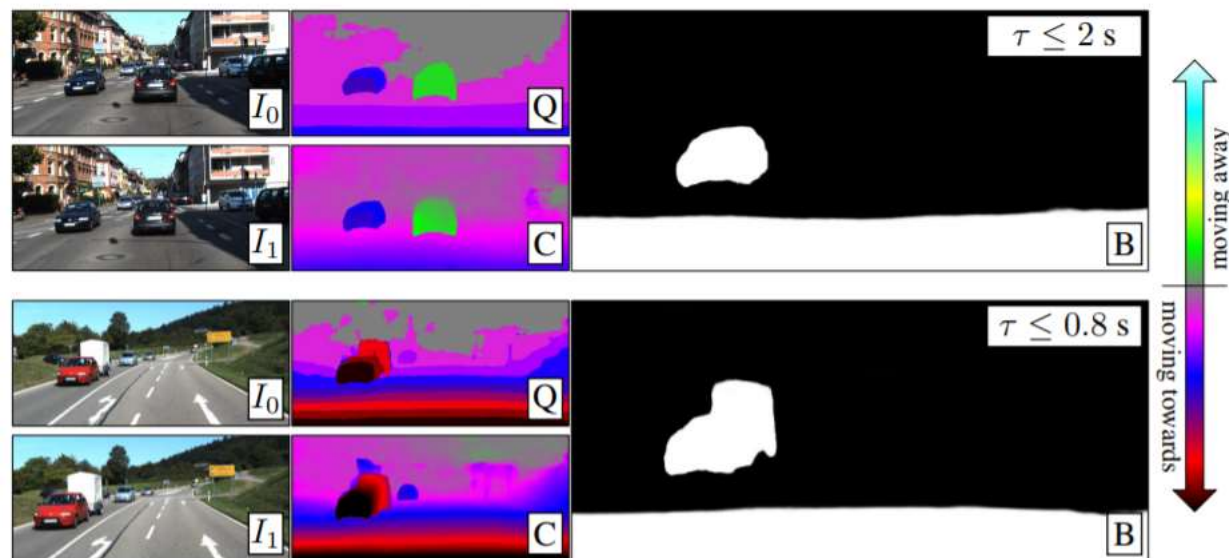

Figure 1: Given I_0 and I_1, our binary time-to-contact (TTC) estimation acts as a temporal geofence detecting objects that will collide with the camera plane within a given time, $\boxed{B}$. It only takes 6.4 ms to compute. Our method can also output quantized TTC, $\boxed{Q}$, and continuous TTC, $\boxed{C}$.
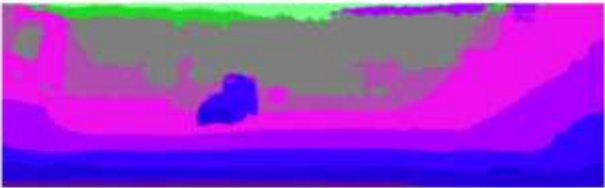
閱讀心得文件與復現紀錄

Binary TTC

Fig. 13. quant-ttc-out.jpg

(i) 測試 - eta-out.png　　　(ii) 測試 - of-out.png

Fig. 14. cont_ttc-of

(i) 測試 - 002-thr0.80-img-in-0.jpg　　(ii) 測試 - 002-thr0.80-img-in-1.jpg　　(iii) 測試 - 002-thr0.80-seg-out.jpg

Fig. 6. 002-thr0.80

(i) 測試 - 003-thr0.85-img-in-0.jpg　　(ii) 測試 - 003-thr0.85-img-in-1.jpg　　(iii) 測試 - 003-thr0.85-seg-out.jpg

Fig. 7. 003-thr0.85

(i) 測試 - 004-thr0.90-img-in-0.jpg　　(ii) 測試 - 004-thr0.90-img-in-1.jpg　　(iii) 測試 - 004-thr0.90-seg-out.jpg

END