

# TrackNet v1

NCTU

TrackNet: maximise  $\partial l(x, y, t) / \partial t$

channel 代表在不同時間下同一個 pixel 座標的預測值  
 $L = || \text{heatmap\_pred} - \text{heatmap\_gt} ||$

TSM: model  $P(\text{action} | f(t-k:t+k))$

channel 代表不同位置 feature (抽象概念, 類似 LML 上下文)  
 $L = \text{CE}(\text{action\_pred}, \text{action\_gt})$

對比 TrackNet, Archana (BG sub)

資料 球和背景重疊, 球被球員遮擋, 球速太快有延遲軌跡

軌跡模式 飛行, 擊球, 彈跳

場地 草地, 紅土, 硬地

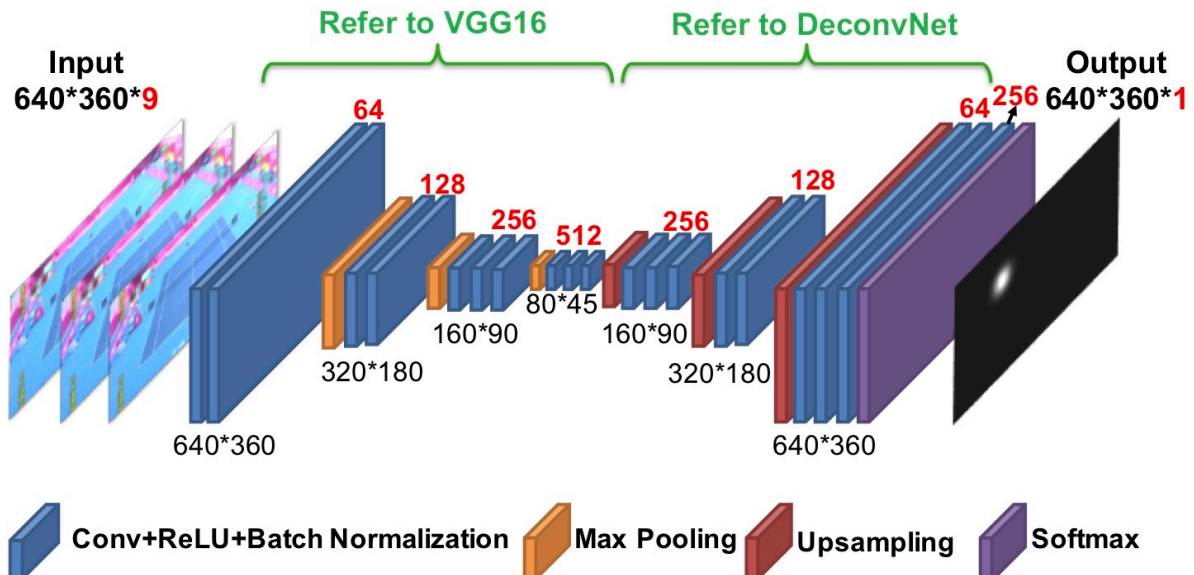
標記 FrameName, Visibility class, X, Y

## 網絡架構

input

$t, t-1, t-2$   
 $640 \times 360 \times 3 \times 3$   $\xrightarrow{\text{RGB}}$

$640 \times 360 \times 9 \rightarrow \text{VGG16} \rightarrow 80 \times 45 \times 512 \rightarrow \text{DeconvNet} \rightarrow 640 \times 360 \times 1$



output heatmap 轉換做單向二元素力圖, 閾值為128

$\geq 128 \rightarrow 255$

$< 128 \rightarrow 0$

↓ = 極化 heatmap

Circle hough → 求值則到球

↓ 有 沒有  
返回圓心

## TrackNet v2

Layer	Details	Output size
input	npz file	512 × 288 × 9
conv2d_1	3 × 3 × 64; relu; batch norm	512 × 288 × 64
conv2d_2	3 × 3 × 64; relu; batch norm	512 × 288 × 64
max_pooling_1	2 × 2 max pool; stride 2	256 × 144 × 64
conv2d_3	3 × 3 × 128; relu; batch norm	256 × 144 × 128
conv2d_4	3 × 3 × 128; relu; batch norm	256 × 144 × 128
max_pooling_2	2 × 2 max pool; stride 2	128 × 72 × 128
conv2d_5	3 × 3 × 256; relu; batch norm	128 × 72 × 256
conv2d_6	3 × 3 × 256; relu; batch norm	128 × 72 × 256
conv2d_7	3 × 3 × 256; relu; batch norm	128 × 72 × 256
max_pooling_3	2 × 2 max pool; stride 2	64 × 36 × 256
conv2d_8	3 × 3 × 512; relu; batch norm	64 × 36 × 512
conv2d_9	3 × 3 × 512; relu; batch norm	64 × 36 × 512
conv2d_10	3 × 3 × 512; relu; batch norm	64 × 36 × 512
up_sampling_1	2 × 2	128 × 72 × 512
concatenate_1	with conv2d_7; axis = 1	128 × 72 × 768
conv2d_11	3 × 3 × 256; relu; batch norm	128 × 72 × 256
conv2d_12	3 × 3 × 256; relu; batch norm	128 × 72 × 256
conv2d_13	3 × 3 × 256; relu; batch norm	128 × 72 × 256
up_sampling_2	2 × 2	256 × 144 × 256
concatenate_2	with conv2d_4; axis = 1	256 × 144 × 384
conv2d_14	3 × 3 × 128; relu; batch norm	256 × 144 × 128
conv2d_15	3 × 3 × 128; relu; batch norm	256 × 144 × 128
up_sampling_3	2 × 2	512 × 288 × 128
concatenate_3	with conv2d_2; axis = 1	512 × 288 × 192
conv2d_16	3 × 3 × 64; relu; batch norm	512 × 288 × 64
conv2d_17	3 × 3 × 64; relu; batch norm	512 × 288 × 64
conv2d_18	1 × 1 × 3; relu	512 × 288 × 3

TABLE 1

TRACKNETV2 MODEL STRUCTURE.

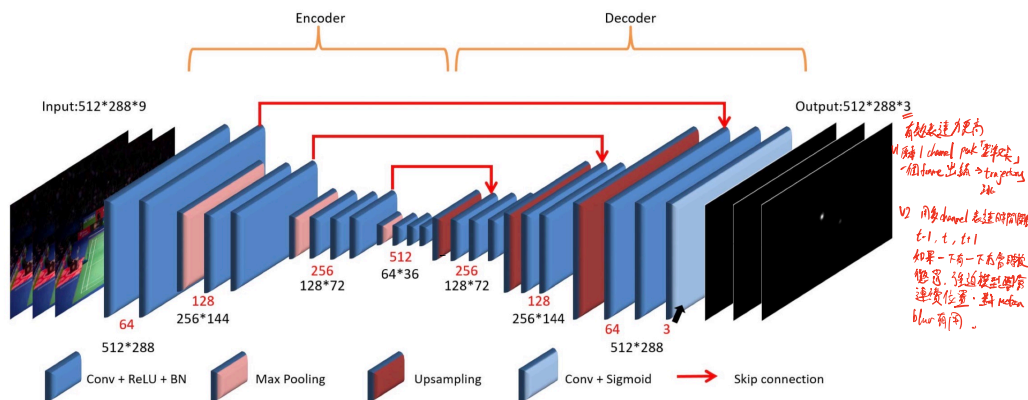


Fig. 2. The architecture of TrackNetV2. The encoder-decoder structure basically follows the design of TrackNet. Compared with the original TrackNet, the innovation includes the downsizing of input images, skip connections enlightened by U-net, concise heatmap representation, and MIMO network design.

V1 GT heatmap 偏向 hard binary

Loss 對 notan blur frame 不穩定 (BCE)

decoder 上採樣太 aggressive

Output 只有 1 channel

BG 樣本少, 少 FP

data imbalance

(Weighted BCE)

V2 Gaussian heatmap 更平滑 (類似 soft label)

□ 根據球 size / resolution 調整

peak 每個一個 pixel

decoder 更 gradual

Output 3 channel

加 BG frame, FP 更敏感

class balance (球 vs 非球)

更長 warm-up

3/7 skip connection

FP 少, Trajectory jitter 少

Training 穩定,

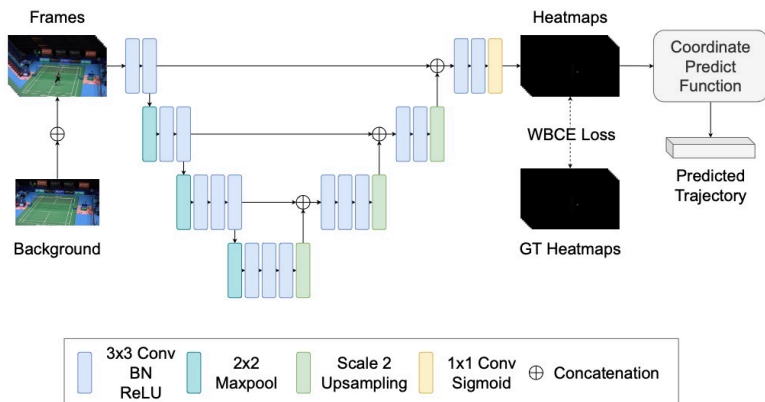
# TrackNet v3

input 中混入背景影像

Training 時使用 Mixup

羽球軌跡校正 (訓練了一個軌跡校正模組) 解決 物體遮蔽、視覺上難以辨識、羽球直接跑出

(a) Tracking module



(b) Rectification module

