## Supplementary materials

Additional information to "Encoder-Decoder Based Convolutional Neural Networks with Multi-Scale-Aware Modules for Crowd Counting" paper. The paper id is 763. The implementation of the paper can be found at

https://github.com/Pongpisit-Thanasutives/Variations-of-SFANet-for-Crowd-Counting

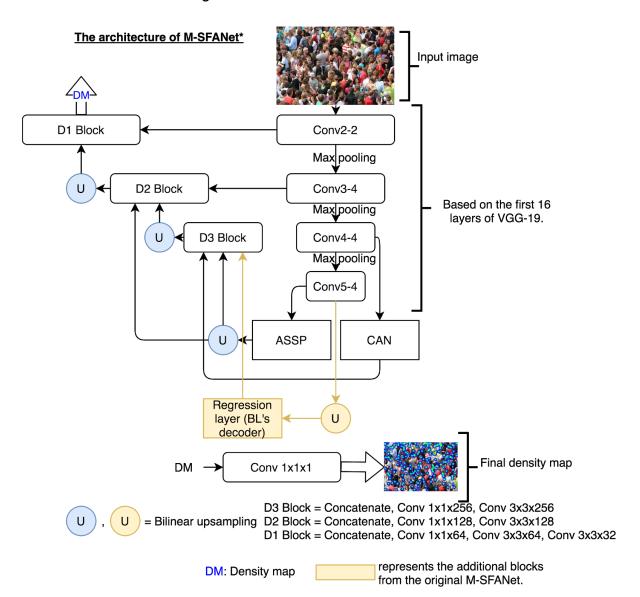


Fig A. The architecture of M-SFANet\*

The architecture of M-SegNet\* Input image FMs D1 Block Conv2-2 Indices Max pooling **FMs** D2 Block Conv3-4 Indices -Max pooling Based on the first 16 FMs D3 Block layers of VGG-19. Conv4-4 Max pooling Regression layer FMs&Indices (BL's decoder) U U Conv5-4 Final density map DM Conv 1x1x1 D3 Block = Concatenate, Conv 1x1x256, Conv 3x3x256 = Max unpooling D2 Block = Concatenate, Conv 1x1x128, Conv 3x3x128 D1 Block = Concatenate, Conv 1x1x64, Conv 3x3x64, Conv 3x3x32 = Bilinear upsampling represents the additional blocks from the original M-SegNet.

Fig B. The architecture of M-SegNet\*

DM: Density map, FMs: Feature maps