Pose Estimation CNN Architecture

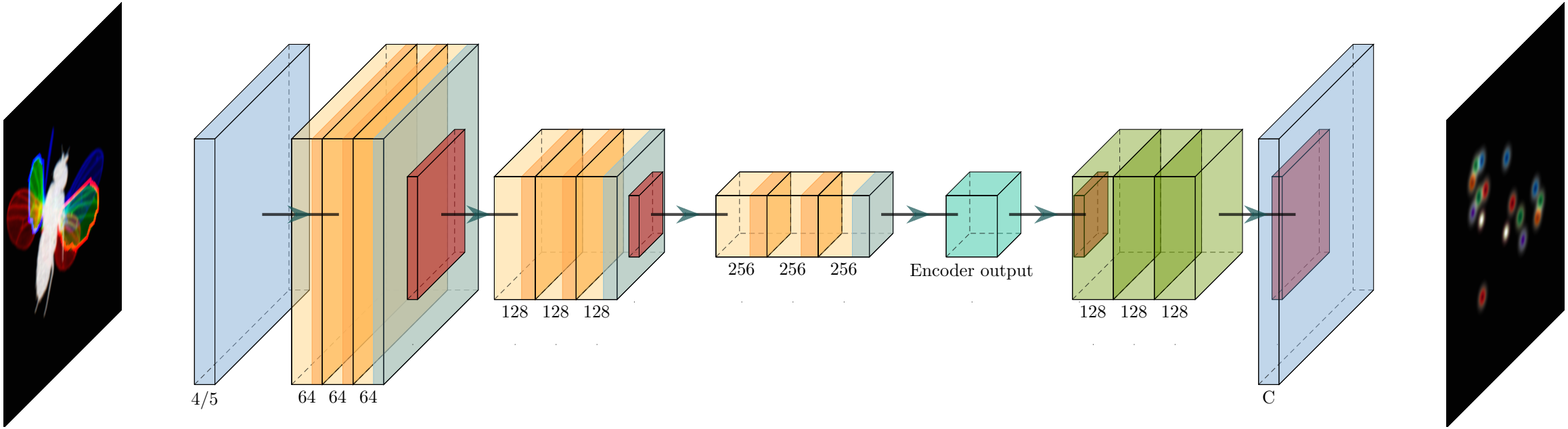

Figure: This architecture is an encoder-decoder design. It processes input images of a fly, enhanced with additional temporal channels and binary masks highlighting the wing locations. These enhancements enable the network to distinguish left and right directions, which are otherwise ambiguous. The network outputs C heatmap channels, where each channel represents a probability density for a specific feature points at each pixel. The feature point for each heatmap is selected as the pixel with the highest intensity value.