Appendix of Paper: Uncertainty-guided Learning for Improving Image Manipulation Detection

Kaixiang Ji$^1$  Feng Chen$^1$  Xin Guo$^1$  Yadong Xu$^2$  Jian Wang$^1$*  Jingdong Chen$^1$

$^1$Ant Group  $^2$Tsinghua University

\{kaixiang.jkx, bangzhu.gx, bobblair.wj, jingdongchen.cjd\}@antgroup.com
chenfeng1271@gmail.com  xuyd17@mails.tsinghua.edu.cn

1. Feature Extractor

It is worth noting that our proposed UEN is applicable to other segmentation-based image manipulation detection methods. Without loss of generality, we present a feature extractor with a general design. Specifically, we adopt HRNetV2 \cite{2} as a basic RGB branch backbone network. The main body of HRNetV2 comprises multiple blocks and is shown in Figure 1. We denote $f_{sr}$ as one block in the $s$-th stage and $r$ is the resolution index. The RGB branch consists of 4 representations: \{f$_{s r}^{\text{RGB}}$; $s \in \{1,2,3,4\}, r \in \{1,2,3,4\}$\}. The input resolution is $\frac{1}{4}$ of the original image resolution because of a preceding stem ahead of $f_{11}$ which comprises two $3 \times 3$ convolutions with stride 2. The resolution of index $r$ is $\frac{1}{2^{r+1}}$ of the original resolution. SRM \cite{3} and resampling feature \cite{1} have shown incredible performance, so we add an SRM branch and a resampling branch parallely. For the SRM branch, we first pass the input image through an SRM layer, accompanied by two consecutive $3 \times 3$ convolution layers with stride 2. Therefore, the SRM branch starts with $\frac{1}{8}$ resolution. The SRM branch consists of 4 representations: \{f$_{sr}^{\text{SRM}}$; $s \in \{1,2,3,4\}, r \in \{1,2,3,4\}$\}. For the resampling branch, following \cite{1}, we extract the resampling feature from $32 \times 32$ non-overlapping patches. Thus, resampling branch starts with $\frac{1}{32}$ resolution, and consists of 1 representation: \{f$_{sr}^{\text{Resampling}}$; $s \in \{4\}, r \in \{4\}$\}. The output embeddings of each branch in the 4-th stage are all concatenated at the corresponding resolution, and the representation head remains consistent with HRNetV2 except for the number of channels.

References


*Corresponding author: bobblair.wj@antgroup.com