Efficient Reference-based Video Super-Resolution (ERVSR): Single Reference Image Is All You Need

Youngrae Kim*, Jinsu Lim*, Hoonhee Cho*, Minji Lee*, Dongman Lee†, Kuk-Jin Yoon†, and Ho-Jin Choi†

1School of Computing, KAIST, 2Mechanical Engineering, KAIST, Daejeon, South Korea
{youngrae.kim, jln2u, gnsngsgml, haewon.lee, kjyoon, hojin}@kaist.ac.kr, dlee@cs.kaist.ac.kr

Abstract

Reference-based video super-resolution (RefVSR) is a promising domain of super-resolution that recovers high-frequency textures of a video using reference video. The multiple cameras with different focal lengths in mobile devices aid recent works in RefVSR, which aim to super-resolve a low-resolution ultra-wide video by utilizing wide-angle videos. Previous works in RefVSR used all reference frames of a Ref video at each time step for the super-resolution of low-resolution videos. However, computation on higher-resolution images increases the runtime and memory consumption, hence hinders the practical application of RefVSR. To solve this problem, we propose an Efficient Reference-based Video Super-Resolution (ERVSR) that exploits a single reference frame to super-resolve whole low-resolution video frames. We introduce an attention-based feature align module and an aggregation upsampling module that attends LR features using the correlation between the reference and LR frames. The proposed ERVSR achieves $12\times$ faster speed, $1/4$ memory consumption than previous state-of-the-art RefVSR networks, and competitive performance on the RealMCVSR dataset while using a single reference image.

1. Introduction

Super-resolution (SR) aims to reconstruct a high-resolution (HR) image from a low-resolution (LR) image. To recover the high-frequency details, reference-based super-resolution (RefSR) approaches [2, 20, 22, 25, 27, 30, 32, 33] utilize additional reference images, e.g. web-crawled high-resolution images [20] and image taken from slightly different viewpoints [16]. The super-resolved image, incorporating high-frequency detail from the reference images, is precise and visually pleasing compared to synthesized textures [33]. In addition, since many recent mobile phones are equipped with dual or triple cameras, the demand for RefSR is increasing.

Reference-based video super-resolution (RefVSR) inherits the advantages of both RefSR and Video Super-Resolution (VSR), which leverages the temporal information in videos. Lee et al. [16] presented the dataset for RefVSR, called RealMCVSR dataset, consisting of ultra-wide and wide-angle videos captured with asymmetric multi-cameras in smartphones. As these cameras capture the scene with different field of view (FoV), the wide-angle image of smaller FoV can be used as a Ref image to super-resolve the ultra-wide angle image of larger FoV. To recover the HR video, Lee et al. [16] computes the correlation between every Ref and LR frame. Guided by the additional information, their approach outperforms VSR and RefSR models. However, their best-performing model consumes 19GB of memory and takes up to $16\times$ longer inference time than the other VSR approaches [4, 5]. This large memory consumption and computation time of RefVSR mainly
comes from the computation of correlation between Ref and LR frames at every time step. Here the question arises: Can we achieve competitive super-resolved video results with less memory consumption while utilizing Ref data? In this paper, we show that competitive performance can be obtained by designing efficient and effective modules for the alignment between Ref and LR data. Unlike the existing work, we do not use Ref frames at every time step, but only use a single Ref frame in the center of the window. As shown in Fig. 1, there is a large overlap between consecutive wide-angle Ref frames. From this point of view, we observe that a single Ref image is sufficient as long as such high-frequency information can be used effectively.

In this paper, we introduce an Efficient Reference-based Video Super-Resolution (ERVSR) framework that super-resolves HR videos from LR videos using a single Ref image. We propose an attention-based feature align (AFA) module which aligns the center Ref frame with the LR frame, thus propagating the Ref information without explicit alignment. Furthermore, high-frequency features of the Ref frame are transferred to every low-resolution feature by attention-based aggregation (AA) upsampling. The AFA module and AA upsampling both exploit widely used attention [21] mechanism where reference features are query and LR features are key and value. Benefiting from attention modules that allow the network to fully exploit a single Ref image, ERVSR achieves competitive performance in the RealMCVSR dataset [16] compared to models that use reference images for every time step [11, 30, 17, 16] with faster speed and less GPU memory.

Our contributions are summarized as follows:

• We propose the ERVSR, which is the first work to tackle the large computation of RefVSR. We successfully optimized the accuracy-efficiency trade-off of RefVSR, opening the possibility of using RefVSR on a mobile device.

• We exploit the attention-based similarity computation and fusion in RefVSR. We also show that a single reference image is enough to recover high-frequency details of the entire video.

2. Related Works

2.1. Reference-based Image Super-resolution

Reference-based image super-resolution methods can be categorized into two: texture transfer methods [32, 30, 27] and the methods that exploit alignment [20, 23]. Texture transfer methods usually unfold images to patches and compute the similarity of reference patches for each LR patch. To measure the similarity, RefSR-Net [32], and TTSSR [27] use the inner product between the raw pixels of patches, whereas SRNTT [30] use the inner product between the features of patches. On the other hand, some works [32, 30] concatenate most similar reference patches for each LR patch and fuse the patches using convolution layers.


2.2. Video Super-Resolution

Many VSR methods perform the alignment by estimating the optical flow field between the target frame and neighboring frames [3, 13, 26], then use convolutions or recurrent networks to fuse features from aligned frames [3, 7]. EDVR [24] aligns the features of each frame using deformable convolution, then fuses them via attention mechanism in both spatial and temporal manner. BasicVSR [4] established the usage of the bidirectional propagation scheme in VSR by showing it maximizes the information gathering. IconVSR [4] extends BasicVSR by adding the additional feature extractor and the coupled propagation mechanism, which interconnect the propagation modules to exploit further information in the sequences. BasicVSR++ [5] extends BasicVSR with second-order grid propagation and flow-guided deformable alignment.

2.3. Reference-based Video Super-Resolution

EFENet [31] utilizes the first frame of a high-resolution ground-truth video as a reference to super-resolve a low-resolution (LR) video. ERVSR is different from EFENet in two folds. First, ERVSR only computes the correlation between the single Ref frame and the single LR frame, whereas EFENet computes flow maps between the Ref and every LR frame using a shared flow estimator. Second, EFENet requires the guidance of an HR video, which is impractical in a real-world scenario.

Lee et al. [16] propose a practical setup that captures the Ref video with an asymmetric multi-camera in a smartphone. They follow a bidirectional propagation scheme, with reference alignment [20] and propagation module in each recurrent cell. Then the aligned reference features are fused with temporally aggregated features using convolutional layers.

Our proposed framework ERVSR differs from previous works in two folds. First, ERVSR does not use the Ref frame every time step, but only one frame in the center time step. There is a lot of overlap between the Ref video frames, and we believe it is not necessary to utilize every Ref frames. Second, existing methods use the reference alignment module proposed in [22] to obtain Ref features aligned to LR frames. Instead of patch-wise calculation of
similarity and alignment, which is computationally heavy, ERVSR computes the similarity between LR and Ref feature using learnable projection [21]. Note that ERVSR is the first approach to exploit attention-based similarity computation and fusion in RefVSR.

### 3. Proposed Method

In this section, we present ERVSR, an end-to-end neural network for efficient reference-based video super-resolution. As shown in Fig. 2, our proposed network consists of three components: Attention-based Feature Align (AFA), Bidirectional Propagation, and Attention-based Aggregation (AA) upsampling module. The AFA module aligns the LR frame at center time \( t \), the Bidirectional Propagation propagates the temporal information to the other time steps, and the AA module transfers the features from the Ref image to upsample the LR feature.

The LR frames \( \{ I_{t-w}^{LR}, \ldots, I_t^{LR}, \ldots, I_{t+w}^{LR} \} \in \mathbb{R}^{3 \times W \times H} \) and a reference frame \( I_t^{ref} \) are first deeply represented by residual layers [8], resulting \( F_{LR}^t, F_{ref}^t \in \mathbb{R}^{N_c \times W \times H} \) significantly, where \( N_c \) denotes the number of channels. Since the properties of the features to be extracted from the LR frame and Ref frame are different, the two residual blocks (ResBlock) do not share weights.

#### 3.1. Attention-based Feature Align (AFA) Module

Extracted features are normalized through normalization layer [1] and projected to query \( Q \), key \( K \), and value \( V \) [21]. We calculate the \( Q, K, V \) as follows:

\[
Q = P_Q(F_t^{ref}) \in \mathbb{R}^{WH \times N_c}, \\
K = P_K(F_t^{LR}) \in \mathbb{R}^{N_c \times WH}, \\
V = P_V(F_t^{LR}) \in \mathbb{R}^{WH \times N_c},
\]

where \( P_Q, P_K, P_V \) denote the projection consisting of \( 1 \times 1 \) convolutional layer, efficient depth-wise separate convolutional layer [9], and reshaping operation. AFA module then computes the correlation between projected reference fea-
Figure 4. The overview of bidirectional propagation. The bidirectional propagation consists of cascaded forward (FW) and backward (BW) modules.

Figure 5. The proposed Attention-based Aggregation (AA) upsampling module.

\[
A = \text{Softmax}(KQ) \in \mathbb{R}^{N_c \times N_c}.
\]  

Projected LR feature \( V \) is attended using the information of reference, \( A \), resulting an aligned LR feature \( F_t^{align} \) where out projection \( P_O \) includes \( 1 \times 1 \) convolutional layer and reshaping operation:

\[
F_t^{align} = P_O(VA) + F_t^{LR}.
\]  

In attention calculation, we also utilize depth-wise separate convolutions for efficient computation as demonstrated in Fig. 3. For the forward network, we adopt the Gated-Dconv Feed-forward Network (GDFN) in [28]. The AFA module is repeated \( N \) times. Using attention mechanism, ERVSER can fully exploit Ref features, not only the most similar features, and benefit from repetitive textures.

By extracting and propagating \( F_t^{align} \) instead of the low-resolution feature, we can exploit information from reference efficiently. As proposed in [16], aligning all LR frames with reference frames would guide more information to the network, but it leads to a huge computational cost. Therefore, LR frames other than \( F_t^{LR} \), are extracted to features by residual block, not the AFA module, resulting \( F_i \) where \( i \in \{t - w, \ldots, t + w\} \) and \( i \notin \{t\} \).

### 3.2. Bidirectional Propagation

The generally used approach to propagate the temporal information is the unidirectional propagation method [10], where the information is sequentially propagated from the first frame to the last frame. However, the problem with this method is that the first frame receives no information from the other frames. In other words, there can be an information imbalance. Therefore, instead of a unidirectional propagation approach, we adopted the bidirectional method [4] to propagate temporal information in each frame.

As shown in Fig. 4, given consecutive LR features \( \{F_{t-1}, F_t^{align}, F_{t+1}\} \), we can obtain the forward feature \( h_t^{fw} \) and backward feature \( h_t^{bw} \) where \( fw \) and \( bw \) denote...
4. Experiments

4.1. Dataset

Our model is trained and tested on the RealMCVSR dataset [16]. RealMCVSR dataset provides real-world HD video triplets concurrently recorded by Apple iPhone 12 Pro Max equipped with triple cameras having fixed focal lengths: ultra-wide (30mm), wide-angle (59mm), and telephoto (147mm). The video triplets are split into training, validation, and testing sets, each of which has 137, 8, and 16 triplets of 19,426, 1,141, and 2,540 frames, respectively. Following the [16], we set the ultra-wide and wide-angle as 16 triplets of 19,426, 1,141, and 2,540 frames, respectively.

4.2. Implementation Details

The network and experiments are implemented using the Pytorch framework. We use NVIDIA A100-40GB when measuring the inference time and GPU memory usage. For the training, we use $\ell_1$ loss as a loss function and Adam [14] optimizer. For each iteration, we randomly sample batches of frame triplets from the RealMCVSR training set while setting the batch size 32. We used the pretrained optical flow network [18] in bidirectional propagation. We trained our model to super-resolve a 4× bicubic downsampled LR

<table>
<thead>
<tr>
<th>Reference usage</th>
<th>Model</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
<th>Inference time (ms/frame)</th>
<th>GPU memory usage (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image SR</td>
<td>Bicubic</td>
<td>No</td>
<td>26.65</td>
<td>0.8</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>SRGAN [15]</td>
<td>No</td>
<td>29.38</td>
<td>0.877</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>RCAN-\ell_1 [29]</td>
<td>No</td>
<td>31.07</td>
<td>0.915</td>
<td>-</td>
</tr>
<tr>
<td>Reference Image SR</td>
<td>TTSSR [27]</td>
<td>All time steps</td>
<td>30.31</td>
<td>0.905</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>TTSSR-\ell_1 [27]</td>
<td>All time steps</td>
<td>30.83</td>
<td>0.911</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DCSR [23]</td>
<td>All time steps</td>
<td>30.58</td>
<td>0.887</td>
<td>-</td>
</tr>
<tr>
<td>Video SR</td>
<td>DCSR-\ell_1 [23]</td>
<td>All time steps</td>
<td>30.63</td>
<td>0.895</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DCSR-\ell_1 [23]</td>
<td>All time steps</td>
<td>32.43</td>
<td>0.933</td>
<td>-</td>
</tr>
<tr>
<td>Reference Video SR</td>
<td>Lee et al.-IR [16]</td>
<td>All time steps</td>
<td>31.73</td>
<td>0.916</td>
<td>1204.61</td>
</tr>
<tr>
<td>Video SR</td>
<td>Lee et al.-IR-\ell_1 [16]</td>
<td>All time steps</td>
<td>34.86</td>
<td>0.959</td>
<td>1204.61</td>
</tr>
<tr>
<td></td>
<td>ERVSR (Ours)</td>
<td>1 per window</td>
<td>34.44</td>
<td>0.957</td>
<td>107.02</td>
</tr>
</tbody>
</table>

Table 1. Quantitative evaluation for the 13 frames per window on the RealMCVSR dataset. - indicates that the information is not provided in that paper. The best and top-3 results are highlighted. * denotes the trained by ours.

the forward and backward module respectively:

$$h_{fw}^t = f_{bw}(F^{align}_t, F_{t-1}^w, h_{fw}^{t-1}),$$

$$h_{bw}^t = bw(F^{align}_t, F_{t+1}^w, h_{bw}^{t+1}).$$

Each module exploits the flow-based methods for spatial alignment as:

$$s_t^{fw,bw} = O(F^{align}_t, F_{t\pm1}),$$

$$h_t^{fw,bw} = W(h_t^{fw,bw}, s_t^{fw,bw}),$$

$$h_t^{fw,bw} = R(h_t^{fw,bw})(F^{align}_t, h_t^{fw,bw}),$$

where $O$ and $W$ denote the flow estimation and explicit feature-level warping, respectively. Here, $R(h_t^{fw,bw})$ denotes a stack of residuals for each forward and backward warping.

3.3. Attention-based Aggregation (AA) Upsampling

We propose the AA upsampling module to transfer the high-frequency feature of reference while upsampling the LR feature. As illustrated in Fig. 5, given the forward aggregated feature and backward aggregated feature from bidirectional propagation, the final high-resolution (HR) frames are obtained by:

$$I_t^{HR} = U(h_t^{fw}, h_t^{bw}), \quad i \in \{t - w, \ldots, t + w\},$$

where $U$ denotes the upsampling module consisting of pixel-shuffle [19], attention mechanism using AA attention map obtained by computing correlation of LR feature and reference feature. In the same way as the AFA Module, the attention map is calculated using depth-wise separate convolution layer and then scaled up using pixel shuffle. In contrast to the AFA module, AA upsampling module transfers the information from the Ref feature to LR feature for every time step. The AA module is repeated $M$ times.
Figure 6. Qualitative comparison of our methods with previous works. For clarity, the magnified parts of the resultant images are zoomed. From left to right, (a): LR input, (b): Bicubic interpolation, (c): BasicVSR++ [5], (d): Lee et al.-IR-$\ell_1$ [16], and (e): Ours, respectively.

ultra-wide video using a wide-angle video frame as a Ref image. In the training phase, the ultra-wide LR frames and wide-angle Ref frames are cropped to $64 \times 64$. We set the number of layers to $N = M = 4$ (in sec 3.1, 3.3) equally in our network and the window size to 13 as a default size. More details about the experiment can be found in our supplementary materials.

4.3. Quantitative Comparison

Table [1] shows a quantitative comparison on RealM-CVSR dataset. We used $4 \times$ bicubic downsampled low-resolution ultra-wide video and wide-angle reference video as input. We compared ERVSR with previous works: SRGAN [15], RCAN [29], TTSR [27], DCSR [23], EDVR [24], BasicVSR [4], BasicVSR++ [5], Lee et al. [16]. We demonstrated the comparisons on PSNR, SSIM, inference time per frame, and GPU memory usage. Models trained only on $\ell_1$ loss function are indicated with $\ell_1$.

Our ERVSR outperforms the previous single image SR (SISR), RefSR, and VSR methods and achieves the second-best in both PSNR and SSIM among several models. The performance difference between Lee et al.-IR-$\ell_1$ [16] (best performing model among several models in [16]) and our model is acceptable considering that our model is 12 times faster, consumes 3.8 times smaller GPU memory usage, and uses only one Ref image at inference time. Even when only one Ref image is given, we show that our method efficiently and effectively exploits and transfers the high-frequency information from the Ref image by using our novel attention-based feature align (AFA) module (in sec 3.1) and attention-based aggregation (AA) upsampling module (in sec 3.3). The efficiency of ERVSR is also competitive with the methods in VSR in inference time per frame and GPU memory usage, while PSNR and SSIM values are much higher. This indicates that the efficiency of our method is comparable with the VSR models, even incorporating the Ref frame in
the network. Therefore, our proposed ERVSR successfully reduces the accuracy-efficiency of the trade-off of RefVSR.

### 4.4. Qualitative Comparison

In Fig. 6 we show the super-resolved results of Bicubic, BasicVSR++ [5], and Lee et al.-IR-[ℓ] [16] trained on RealMCVSR dataset. Non-reference-based SR methods such as Bicubic and BasicVSR++ tend to produce blurred textures and letters. Contrastingly, RefVSR methods benefit from the Ref information and better super-resolve the details such as letters. Also, our proposed ERVSR shows competitive visual quality with Lee et al.-IR-[ℓ] [16], even though our method uses only one Ref image per window.

Lee et al.-IR-[ℓ] [16] often produces the blurry artifact, especially in small-sized features and non-overlapping FoV area. It is due to the misalignment between the LR image and the Ref image since it explicitly calculates the similarity map. Benefiting from the attention-based mechanism, our ERVSR can provide better visual results on small-sized features by utilizing the feature-level alignment. Also, since our AA module transfers the high-resolution detail to LR from the Ref feature for overlapping as well as non-overlapping regions, ERVSR can make better results in the non-overlapping FOV area.

### 4.5. Ablation Studies

1. **Contribution of each component of ERVSR to the performance.** We conducted quantitative ablation studies to analyze the effect of each of our proposed modules, AFA and AA. We set the baseline network as a model solely consisting of bidirectional propagation without AFA and AA modules. As demonstrated in Table 2, the models with our each proposed module show better PSNR and SSIM than the baseline network. AA module has a more remarkable performance improvement than the AFA module since it is effective in directly transferring the LR feature to temporally aligned features by bidirectional propagation. We find that the performance gain is significant when the network uses both AA and AFA modules. We hypothesize that the network needs to learn the encoded and decoded feature simultaneously. Using the solely consisted module becomes extremely difficult to transfer the high-resolution information from the Ref feature to the LR feature. On the other hand, using both AA and AFA module at encode- and decode-level achieve to align the Ref to LR feature without explicit alignment.

As shown in Fig. 7, we also conducted qualitative ablation studies to show the visual effect of each component on SR. The Baseline cannot restore the details such as letters or repetitive patterns. The Baseline + AFA shows reduced distortion than the Baseline but still produces blurry artifacts. Our ERVSR model with both AFA and AA modules achieves the best visual quality with reduced blurry artifacts and distortion and accurate restoration of edges.

2. **Can reference at the center frame guide the LR frames with a large time difference?** This question is crucial since we use a single reference frame per window. To answer this question, we compared the PSNR of each super-resolved video frame of ERVSR and BasicVSR++ with respect to its time difference from the center frame (See Fig. 8). Performance degradation decreases as the time difference increases, but such an amount of degradation is also observed in BasicVSR++. If reference at the center frame cannot effectively guide the LR frames with a large time difference, the performance degradation will be greater than that of the VSR. However, the performance gap between ERVSR and BasicVSR++ is constant, which implies that the details of the reference center frame manage to guide the feature extraction and upsampling of an LR frame even with a large time interval.

3. **Effect of frame number in performance.** VSR models can benefit from the increased number of input LR frames since the feature can be better refined from the bidirectional flow. The advantageous effect of increased frame number escalates in RefVSR models that exploit reference frames for every time step. In this section, we show the comparison of PSNR and SSIM for the number of frames in the window in Table 3. As the frame number decrease, the performance gap between our model and Lee et al.-IR-[ℓ] [16] decreases, and finally, our model outperforms it when the frame number is 5. Knowing that Lee et al.-IR-[ℓ] [16] still exploits four more reference frames than ERVSR when the frame number is 5, such results imply that our attention-based modules are effective design for the utilization of the reference details.

### 5. Conclusion

In this work, we propose efficient reference-based video super-resolution (ERVSR) using only a single reference image. To this end, we first propose the attention-based approach to compute similarity using learnable projections in RefVSR. To evaluate our method, we train and evaluate our network in a large-scale benchmark in [16]. Our model outperforms the state-of-the-art image SR, reference-based image SR, and VSR methods in both qualitatively and quantitatively. We accelerate the inference time for \( \times12 \) faster while using GPU memory to \( \times3.8 \) efficiently.

<table>
<thead>
<tr>
<th>Model</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>31.88</td>
<td>0.931</td>
</tr>
<tr>
<td>Baseline + AFA</td>
<td>32.11</td>
<td>0.932</td>
</tr>
<tr>
<td>Baseline + AA</td>
<td>32.17</td>
<td>0.933</td>
</tr>
<tr>
<td>ERVSR (Baseline + AFA + AA)</td>
<td>34.44</td>
<td>0.957</td>
</tr>
</tbody>
</table>

Table 2. Quantitative ablation study of our proposed components.
Figure 7. Qualitative results of ablation studies on each component. For clarity, the magnified parts of the resultant images are zoomed in. From left to right, (a): LR input, (b): Baseline, (c): Baseline + AFA, (d): Baseline + AA, and (e): ERVSR (Baseline + AFA + AA), respectively.

Figure 8. Comparison on PSNR of each video frame.

Table 3. Quantitative evaluation of various networks on various the number of frames in window. Best results are highlighted and second-best results are underlined. 1st and 2nd row mean PSNR (db) and SSIM scores, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>The number of frame for SR</th>
<th>5</th>
<th>7</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>BasicVSR++</td>
<td>32.56</td>
<td>32.74</td>
<td>32.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9381</td>
<td>0.9404</td>
<td>0.9416</td>
<td></td>
</tr>
<tr>
<td>Lee et al.-IR-ℓ₁ [16]</td>
<td>34.02</td>
<td>34.36</td>
<td>34.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9516</td>
<td>0.9548</td>
<td>0.959</td>
<td></td>
</tr>
<tr>
<td>ERVSR (Ours)</td>
<td>34.03</td>
<td>34.15</td>
<td>34.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9534</td>
<td>0.9541</td>
<td>0.9567</td>
<td></td>
</tr>
</tbody>
</table>

where only one reference frame in a video is used, compared with Lee et al.-IR-ℓ₁ [16], which is the state-of-the-art approach in RefVSR domain. In short, we reduce the accuracy-efficiency trade-off of RefVSR, opening the possibility of using RefVSR in real-time problems.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01126, Self-learning based Autonomic IoT Edge Computing). (50%)
This research was supported and funded by the Korean National Police Agency. [Project Name: XR Counter-Terrorism Education and Training Test Bed Establishment / Project Number: PR08-04-000-21] (50%)

References


