STREAMLINED HYBRID ANNOTATION FRAMEWORK USING SCALABLE CODESTREAM FOR BANDWIDTH-RESTRICTED UAV OBJECT DETECTION

Karim El Khoury[⋆]*, Tiffanie Godelaine*[⋆]*, Simon Delvaux, Sébastien Lugan, and Benoît Macq*

ICTEAM, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

ABSTRACT

Emergency response missions depend on the fast relay of visual information, a task to which unmanned aerial vehicles are well adapted. However, the effective use of unmanned aerial vehicles is often compromised by bandwidth limitations that impede fast data transmission, thereby delaying the quick decision-making necessary in emergency situations. To address these challenges, this paper presents a streamlined hybrid annotation framework that utilizes the JPEG 2000 compression algorithm to facilitate object detection under limited bandwidth. The proposed framework employs a fine-tuned deep learning network for initial image annotation at lower resolutions and uses JPEG 2000's scalable codestream to selectively enhance the image resolution in critical areas that require human expert annotation. We show that our proposed hybrid framework reduces the response time by a factor of 34 in emergency situations compared to a baseline approach.

Index Terms— Scalable Codestream, JPEG 2000, UAV, Object Detection, Hybrid Annotation

1. INTRODUCTION

The utilization of Unmanned Aerial Vehicles (UAVs) has become standard practice in improving the operational capabilities of first responders in emergency situations. UAVs provide essential real-time visual information for areas affected by disasters, aiding in damage assessment, hazard detection, and search and rescue missions [1, 2]. Hybrid annotation stands out as a method to improve the accuracy and reliability of data interpretation in such challenging scenarios. Hybrid annotation begins with an automated Deep Learning (DL)-based object detection, which is then refined by human annotators in areas of uncertainty. The aim is to enhance the annotation process to accomplish two essential objectives: (1) allowing first responders to make fast and accurate decisions, and (2) saving human expert annotators' valuable time. However, the effectiveness of conventional hybrid annotation frameworks depends on the reliability of data transmission. In emergency scenarios, limited bandwidth, obstructed line of sight and damaged infrastructure delay UAV-to-ground data transmission, hindering the decision-making process.

[⋆]The authors have contributed equally to this work.

In this paper, we put forward a streamlined hybrid annotation framework using scalable codestream for bandwidth-restricted UAV object detection. Our framework incorporates the JPEG 2000 (J2K) compression algorithm within a hybrid annotation workflow. The scalability of the J2K codestream allows for multiresolution access to specific regions of interest, particularly useful when dealing with limited bandwidth [3]. The proposed approach streamlines the hybrid annotation process by initially analyzing low-resolution images through a fine-tuned DL network. For areas requiring more detailed analysis, the J2K algorithm facilitates selective enhancement at high resolution, enabling human experts to perform precise annotations.

The contributions of our work are summarized as follows:

- We propose an optimized hybrid annotation framework that takes into account the scenario-specific bandwidth constraints to ensure a fast decision-making process. We observe that the proposed framework reduces the response time by a factor of 34 compared to a baseline approach.
- We show that fine-tuning a DL network for object detection on low-resolution images remains effective while providing fast information access in bandwidth-restricted situations.

2. RELATED WORKS

In order to categorize the contribution of each of the related works, we define the following four research challenges: (I) High-resolution image compression, (II) DL-based object detection, (III) Hybrid annotation, and (IV) Applicability to bandwidth-restricted UAV scenarios. Table 1 provides a summarized comparison of the related works as per the aforementioned challenges. We focus on contributions that address at least two of the four challenges. To the best of our knowledge, the proposed streamlined hybrid annotation framework using scalable codestream for bandwidth-restricted UAV object detection is the first contribution to offer a comprehensive solution to all four challenges.

2.1. Related works merging (I) and (IV):

Zhou et al. [4] present a comprehensive review of remote sensing image compression algorithms. They highlight the

J2K codec's ability to achieve high compression ratios without significant loss of image quality which is crucial for efficient transmission of high-resolution images in limited bandwidth scenarios. Indradjad et al. [5] compare four wavelet-based methods for compressing satellite imagery. They show that J2K delivers better image quality than other wavelet-based codecs at lower bit rates.

2.2. Related works merging (I) and (II):

Ehrlich et al. [6] introduce a self-supervised artifact correction approach that improves object detection accuracy on JPEG compressed images. Ghazvinian Zanjani et al. [7] establish that DL models can accurately identify metastatic cancer in J2K compressed histopathological images. El Khoury et al. [8, 9] show that fine-tuning a 3D U-Net improves its ability to handle J2K compression artifacts while maintaining precise organ detection and segmentation on radiotherapy scans.

2.3. Related works merging (I), (II) and (IV):

Zhang et al. [10] develop a DL-based mechanism for UAV-aided networks that compresses media data into constant-sized latent codes for efficient real-time transmission, successfully maintaining the quality of visual data for small objects despite bandwidth constraints and compression artifacts. However, their proposed system relies on having uninterrupted line-of-sight communication which limits its applicability in obstructed and complex environments. Preethy Byju et al. [11] design a DL framework for efficiently classifying scenes in J2K compressed remote-sensing images, which balances computational efficiency with detection accuracy. Given that they work on UAV scenarios with large compressed data repositories, their contribution is not tailored for real-time use. Both contributions do not work within a hybrid annotation framework.

2.4. Related works merging (II), (III) and (IV):

Kellenberger et al. [12] present an active learning method that requires limited labeled data from human annotators to improve animal detection in UAV imagery. Yamani et al. [13] introduce a single-stage object detection active learning framework for UAV imagery that minimizes annotation expenses through an innovative uncertainty aggregation technique while mitigating class imbalance. Both contributions do not consider the image compression requirements imposed by low-bandwidth constraints.

3. PROPOSED APPROACH

In this section, we first present a conventional baseline hybrid annotation framework and then detail our proposed streamlined hybrid annotation framework. It should be noted that both frameworks are based on the J2K scalable codestream for encoding and decoding as well as multiresolution tile extraction.

Table 1. Comparison of related works with respect to four research challenges: (I) High-resolution image compression, (II) DL-based object detection, (III) Hybrid annotation, and (IV) Applicability to bandwidth-restricted UAV scenarios.

	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	Ours
(I)	✓	✓	✓	✓	✓	✓	✓	✓			✓
(II)			✓	✓	✓	✓	✓	✓	✓	✓	✓
(III)									✓	✓	✓
(IV)	✓	✓					✓	✓	✓	✓	✓

3.1. Baseline hybrid annotation framework

The baseline framework aims to emulate the standard hybrid annotation approach used by UAVs in emergency situations. Figure 1 illustrates a simplified block diagram of the framework, modeling a conventional compression scheme through a lossless UAV-to-ground station communication channel. A step-by-step pseudo-code of the baseline hybrid annotation framework is presented in Algorithm 1.

The basic working principles of the baseline framework are described as follows. The framework takes as input a high-resolution still image (I^{HR}), a predefined high-resolution level (HR), the high-resolution tile size for the J2K encoding ($\texttt{SIZE}(T^{HR})$), and a predetermined human annotation budget ($|\mathscr{N}|$), which represents how many tiles a human annotator can afford to annotate in time-limited situations. The first step involves calculating the total number of tiles ($|\mathscr{M}|$) in I^{HR}. The high-resolution tiles (T^{HR}) from I^{HR} are then encoded using J2K and sent via a lossless communication channel to the ground station. The tiles are then decoded and processed by a DL model to produce preliminary annotations. The *indexer* block, detailed in Subroutine 1, processes the resultant annotations, made up of the bounding boxes (BB_{DL}) and their confidence scores (CS_{DL}). It identifies a set of tiles that require inspection by a human expert annotator and stores their indices ($\mathscr{N}$). The framework then calls the *extractor* block, detailed in Subroutine 2, to process the J2K scalable codestream and extracts the specific $\mathscr{N}$ tiles in high resolution. The $\mathscr{N}$ selected tiles (T_n^{HR}) are then annotated by the human expert. The resultant bounding boxes (BB_{HUM}) and confidence scores (CS_{HUM}) are sent out along with the DL annotations (BB_{DL}, CS_{DL}) to the decision-making unit where critical decisions are taken to respond to emergency situations.

Essentially, the main limitation of the baseline framework is due to its constant transmission of high-resolution tiles from the UAV to the ground station, regardless of the scenario-specific constraints. This becomes problematic when data rates and time demands are low, thereby impeding the framework's effectiveness in time-sensitive situations.

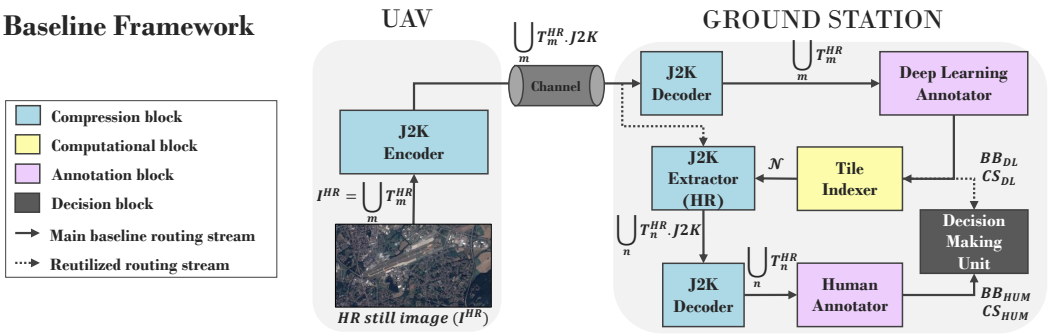

Fig. 1. Simplified block diagram of the baseline hybrid annotation framework.

Algorithm 1: Baseline framework

Require: I^{HR}, HR, SIZE(T^{HR}), $|\mathscr{N}|$
Function main():
- ▷ **Step 0:** Get number of tiles
  $|\mathscr{M}| \leftarrow$ SIZE(I^{HR})/SIZE(T^{HR})
- ▷ **Step 1:** Encode all tiles using J2K
  $\bigcup_m T_m^{HR}.J2K \leftarrow$ J2K(I^{HR}, $\mathscr{M}$)
  where $\mathscr{M} \coloneqq \{1, ..., |\mathscr{M}|\}$
- ▷ **Step 2:** Send tiles via channel
  $\bigcup_m T_m^{HR}.J2K \leftarrow$ TR($\bigcup_m T_m^{HR}.J2K$)
- ▷ **Step 3:** Decode all tiles using J2K
  $\bigcup_m T_m^{HR} \leftarrow$ J2K^{-1}($\bigcup_m T_m^{HR}.J2K$, $\mathscr{M}$)
- ▷ **Step 4:** Annotate high-resolution tiles using DL model
  $BB_{DL}, CS_{DL} \leftarrow$ DL($\bigcup_m T_m^{HR}$, $\mathscr{M}$, HR)
- ▷ **Step 5:** Spot indices of tiles that need human annotation
  $\mathscr{N} \leftarrow$ IND(BB_{DL}, CS_{DL}, $|\mathscr{N}|$)
- ▷ **Step 6:** Extract chosen tiles from J2K
  $\bigcup_n T_n^{HR}.J2K \leftarrow$ EXT($\bigcup_m T_m^{HR}.J2K$, $\mathscr{N}$, HR)
- ▷ **Step 7:** Decode chosen high-resolution tiles using J2K
  $\bigcup_n T_n^{HR} \leftarrow$ J2K^{-1}($\bigcup_n T_n^{HR}.J2K$, $\mathscr{N}$)
- ▷ **Step 8:** Annotate chosen tiles by human annotator
  $BB_{HUM}, CS_{HUM} \leftarrow$ HUM($\bigcup_n T_n^{HR}$, $\mathscr{N}$)
- ▷ **Step 9:** Make decision based on hybrid annotations
  Decision $\leftarrow BB_{HUM}, CS_{HUM}, BB_{DL}, CS_{DL}$
  return *Decision*

3.2. Proposed streamlined hybrid annotation framework

The objective of the proposed framework is to ensure that the decision-making process is rapid and reliable, meeting the critical needs of emergency response operations. To achieve this objective, we propose to modify the baseline framework by streamlining the hybrid annotation process to reduce the overall bandwidth consumption of the system. Figure 2 presents a simplified block diagram of the proposed streamlined hybrid annotation framework, highlighting the main differences with the baseline framework. A step-by-step pseudo-code of the proposed framework is presented in Algorithm 2. The streamlining process is achieved by implementing four improvements to the baseline framework:

Subroutine 1: Indexer

Function IND(BB, CS, $|\mathscr{Y}|$):
- Sort BB w.r.t. CS values in ascending order
- $\mathscr{X}, i \leftarrow \emptyset, 0$
- **while** $i < |\mathscr{Y}|$ **do**
  - $\mathscr{X}$.append($BB[i]$)
  - $i \leftarrow i + 1$
- $\mathscr{Y} \leftarrow \mathscr{X}$
- **return** $\mathscr{Y}$

Subroutine 2: Extractor

Function EXT($\bigcup_x T_x^{HR}.J2K$, $\mathscr{Y}$, R):
- **foreach** $y \in \mathscr{Y}$ **do**
  - Extract T_y^R from $\bigcup_x T_x^{HR}.J2K$ at resolution R
- **return** $\bigcup_y T_y^R.J2K$

1. We define a *budget calculator* block, detailed in Subroutine 3, that takes into account scenario-specific inputs, the transmission time limit ($t_{TRlimit}$) and available data rate ($data\_rate$) from the UAV, as well as the predetermined resolution levels ($Rlvls$) of the J2K encoding. The *budget calculator* block then determines the highest resolution available (LR) at which we can send our images and sets the human annotation budget ($|\mathscr{N}|$) with the remaining bandwidth budget.
2. We fine-tune a DL model on low-resolution images to maintain optimal detection performance at the initial annotation step (BB_{DL}, CS_{DL}).
3. We exploit the J2K scalable codestream by utilizing the tile extraction process at the UAV level, instead of at the ground station level, for both low-resolution and high-resolution tile extraction.
4. We reroute the framework to ensure that only the tiles selected for human annotation are sent in high-resolution, effectively reducing the overall bandwidth consumption of the framework.

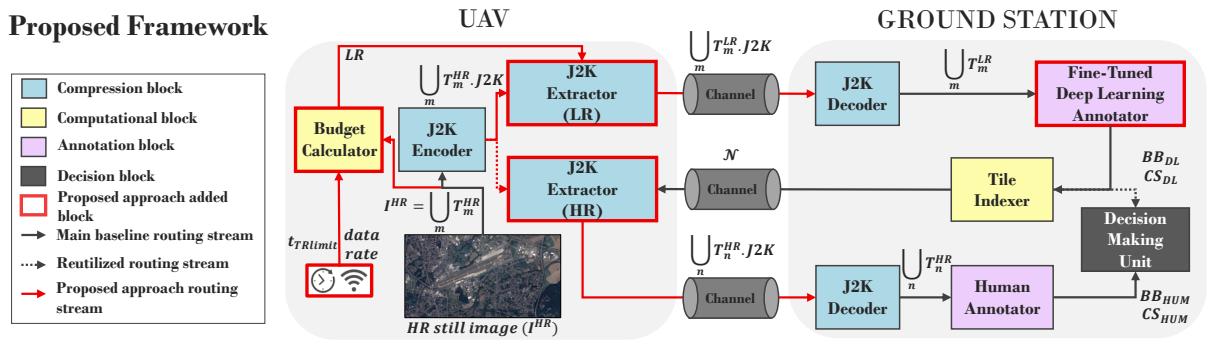

Fig. 2. Simplified block diagram of the proposed streamlined hybrid annotation framework.

Algorithm 2: Proposed streamlined framework

Require: I^{HR}, SIZE (T^{HR}), $data\_rate$, $t_{TRlimit}$, $Rlvls$

Function main **()**:

▷ **Step 0:** Get number of tiles and chosen low resolution
$LR, HR, |\mathcal{N}|, |\mathcal{M}| \leftarrow$ BC $(I^{HR}$, SIZE (T^{HR}), $data\_rate, t_{TRlimit}, Rlvls)$

▷ **Step 1:** Encode all tiles using J2K
$\bigcup_m T_m^{HR}.J2K \leftarrow$ J2K $(I^{HR}, \mathcal{M})$
where $\mathcal{M} := \{1, ..., |\mathcal{M}|\}$

▷ **Step 2:** Extract all tiles in low resolution from J2K
$\bigcup_m T_m^{LR}.J2K \leftarrow$ EXT $(\bigcup_m T_m^{HR}.J2K, \mathcal{M}, LR)$

▷ **Step 3:** Send low-resolution tiles via channel
$\bigcup_m T_m^{LR}.J2K \leftarrow$ TR $(\bigcup_m T_m^{LR}.J2K)$

▷ **Step 4:** Decode all low-resolution tiles using J2K
$\bigcup_m T_m^{LR} \leftarrow$ J2K^{-1} $(\bigcup_m T_m^{LR}.J2K, \mathcal{M})$

▷ **Step 5:** Annotate low-resolution tiles using DL model
$BB_{DL}, CS_{DL} \leftarrow$ DL $(\bigcup_m T_m^{LR}, \mathcal{M}, LR)$

▷ **Step 6:** Spot indices of tiles that need human annotation
$\mathcal{N} \leftarrow$ IND $(BB_{DL}, CS_{DL}, |\mathcal{N}|)$

▷ **Step 7:** Send chosen tile indices via channel
$\mathcal{N} \leftarrow$ TR $(\mathcal{N})$

▷ **Step 8:** Extract chosen tiles in high resolution from J2K
$\bigcup_n T_n^{HR}.J2K \leftarrow$ EXT $(\bigcup_m T_m^{HR}.J2K, \mathcal{N}, HR)$

▷ **Step 9:** Send chosen high-resolution tiles via channel
$\bigcup_n T_n^{HR}.J2K \leftarrow$ TR $(\bigcup_n T_n^{HR}.J2K)$

▷ **Step 10:** Decode chosen high-resolution tiles using J2K
$\bigcup_n T_n^{HR} \leftarrow$ J2K^{-1} $(\bigcup_n T_n^{HR}.J2K, \mathcal{N})$

▷ **Step 11:** Annotate chosen tiles by human annotator
$BB_{HUM}, CS_{HUM} \leftarrow$ HUM $(\bigcup_n T_n^{HR}, \mathcal{N})$

▷ **Step 12:** Make decision based on hybrid annotations
$Decision \leftarrow BB_{HUM}, CS_{HUM}, BB_{DL}, CS_{DL}$

return *Decision*

In summary, the proposed framework utilizes J2K's scalable codestream to reduce the overall bandwidth consumption by first sending a low-resolution image for initial DL network annotation and then transmitting only the selected high-resolution tiles for human expert annotation.

Subroutine 3: Budget calculator

Function BC $(I^{HR}$, SIZE (T^{HR}), $data\_rate$, $t_{TRlimit}$, $Rlvls)$:

$LR \leftarrow -1$
$|\mathcal{N}|, |\mathcal{M}| \leftarrow 0$
$BW \leftarrow data\_rate \cdot t_{TRlimit}$
$HR \leftarrow$ MAX $(Rlvls)$
for $R \leftarrow HR$ **to** 1 **do**
  if SIZE $(I^R) \leq BW$ **then**
    $LR \leftarrow R$
    $|\mathcal{N}| \leftarrow (BW -$ SIZE $(I^R))$ / SIZE (T^{HR})
    break
if $LR = -1$ **then**
  Insufficient BW Budget, must relax constraints
$|\mathcal{M}| \leftarrow$ SIZE (I^{HR}) / SIZE (T^{HR})
return $LR, HR, |\mathcal{N}|, |\mathcal{M}|$

4. EXPERIMENTAL SETUP

In this section, we introduce the materials used, the metrics applied and the experiments done to evaluate the proposed framework performance.

4.1. Materials

The dataset used is composed of 136 high-resolution satellite still images representing airports in different locations [14]. We train a YOLOv8[1] model for multi-class identification of civilian aircraft. Five models are fine-tuned on five resolution levels with a training/validation/test sets split of 89/42/5 images, respectively. The J2K encoding, extraction and decoding tasks, are performed using the OpenJPEG platform[2].

4.2. Evaluation metrics

Two evaluation metrics are used to compare the baseline and the proposed frameworks: **response time ratio** (t_{RS}_ratio) and **recall difference** ($recall\_diff$).

[1]https://github.com/ultralytics/ultralytics
[2]https://www.openjpeg.org/

t_{RS}_ratio is used to compare the response times between the proposed and the baseline frameworks. For each framework, the response time (t_{RS}), the time at which the decision is taken, is defined as:

$$t_{RS} = t_{TR} + t_{HUM} \tag{1}$$

where t_{TR} is the transmission time and t_{HUM} is the human annotator time.

The transmission time (t_{TR}) is the time needed for data to be transmitted through the channel. We distinguish between the transmission times of the two frameworks as follows.

The baseline framework transmission time is defined as:

$$t_{TR} = \frac{\texttt{SIZE}(\bigcup_m T_m^{HR}.\textit{J2K})}{data\_rate} \tag{2}$$

where $data\_rate$ varies depending on the scenario at hand.

The proposed framework transmission time is defined as:

$$t_{TR} = \frac{\texttt{SIZE}(\bigcup_m T_m^{LR}.\textit{J2K}) + \texttt{SIZE}(\bigcup_n T_n^{HR}.\textit{J2K})}{data\_rate} \tag{3}$$

where the framework adapts the tile resolution (LR) depending on the $data\_rate$ and transmission time limit $t_{TRlimit}$ of the scenario at hand.

It should be noted that, for both frameworks, we consider that any additional transmitted metadata is minor in size compared to the image data and therefore has a negligible impact on the total data size.

The human annotator time (t_{HUM}), the time required for human experts to annotate uncertain areas of an image, is defined as:

$$t_{HUM} = \mu_{t_{\text{HUM}}} \cdot |\mathcal{N}| \tag{4}$$

where $\mu_{t_{\text{HUM}}}$ is the average time a human expert needs to annotate a single tile and $|\mathcal{N}|$ is the number of tiles requiring expert annotation.

The **response time ratio** (t_{RS}_ratio) is thus defined as:

$$t_{RS}\_ratio = \frac{t_{RS}(Baseline)}{t_{RS}(Proposed)} \tag{5}$$

$recall\_diff$ quantifies the difference in object detection performance between the proposed and the baseline frameworks. We consider the $recall$ as the metric for detection accuracy because, in an emergency situation, the priority is to ensure that we have as few false negatives as possible to avoid missing objects of interest.

The $recall$ is defined as:

$$recall = \frac{TP}{TP + FN} \tag{6}$$

where TP is the number of true positive and FN is the number of false negative, a detection being considered as a TP if the intersection over union with the ground truth is higher than 0.1 to prevent missing true positive at low resolution, objects being represented by only few pixels.

The **recall difference** ($recall\_diff$) is thus defined as:

$$\begin{aligned} recall\_diff = {} & recall(t_{RS}(Baseline)) \\ & - recall(t_{RS}(Proposed)) \end{aligned} \tag{7}$$

4.3. Experiments

The objective of the proposed framework is to decrease the response time in emergency scenarios compared to a conventional baseline approach while maintaining optimal object detection. The use case chosen is a surveillance scenario dedicated to the identification of civilian aircraft in multiple airport settings. We configure the J2K compression with five resolution levels and also fine-tune the YOLOv8 model on the same five resolution levels. We test the frameworks on high-resolution images of 60 MP and examine nine emergency scenarios whilst varying two scenario-specific parameters: $t_{TRlimit}$ dependent on the urgency of the emergency response, and the available $data\_rate$. We fix three $t_{TRlimit}$ values of 3, 10, and 30 minutes to emulate three levels of urgency in emergency response scenarios. We consider three $data\_rate$ values: 22, 88, and 176 kbps. These rates fall within the minimum-to-median range available for the Iridium Certus satellite network, the standard for low-bandwidth UAV communications[3]. We do not consider the high data rate range offered by the Iridium Certus network, as we would obtain equivalent performances between our framework and the baseline. We fix $\mu_{t_{\text{HUM}}}$ at 0.5 minutes/tile.

5. RESULTS AND DISCUSSION

The comparison of the proposed and the baseline frameworks for the nine scenarios previously mentioned can be seen in Figure 3. We start by analyzing the least constrained case with a $data\_rate$ of 176 kbps (Figure 3a). When considering $t_{TRlimit}$ values of 3, 10 and 30 minutes, we respectively obtain a penalizing $recall\_diff$ of 0.257, 0.112 and 0.028 while observing a gain in t_{RS}_ratio of 7.08, 3.67 and 1.54. Indeed, there is a trade-off between the t_{RS}_ratio and the $recall\_diff$ when comparing the baseline and the proposed approaches. This trade-off is a direct consequence of the proposed framework's streamlining approach. Given that

[3]https://www.iridium.com/iridiumcertus/

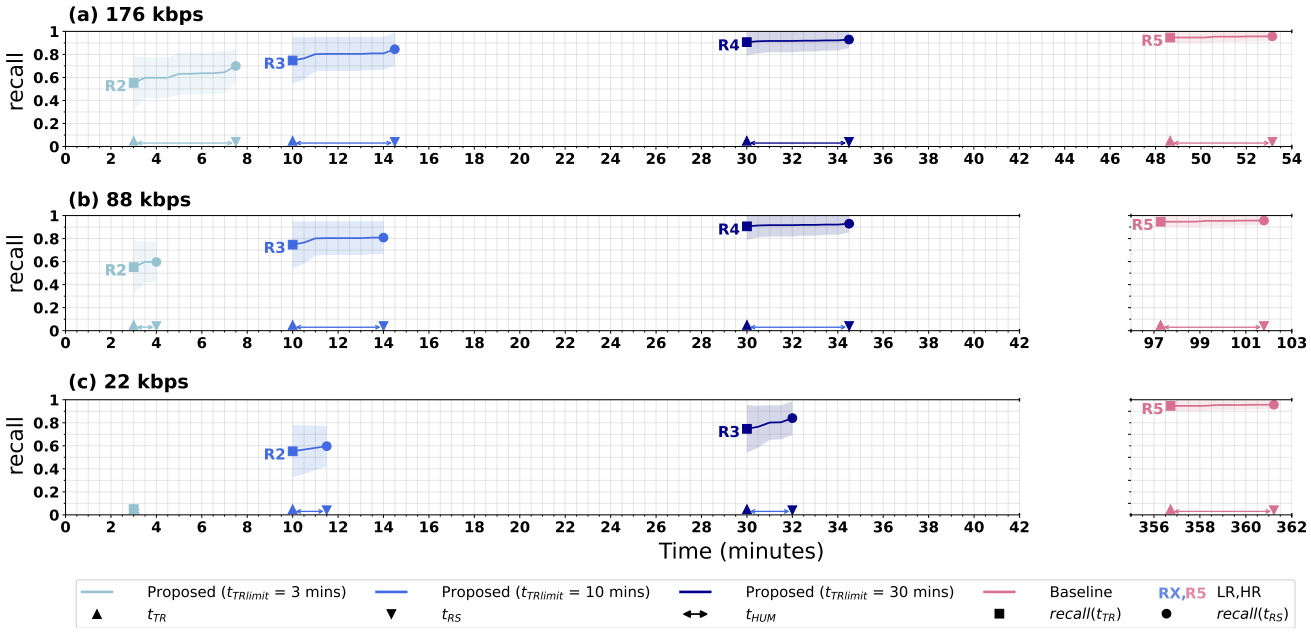

Fig. 3. Recall versus time plot for the baseline framework and the proposed framework for high-resolution image size of 60 MP under a data rate limit of (a) 176 kbps (b) 88 kbps and (c) 22 kbps, with a maximum human annotator time of 5 minutes and considering emergency scenarios with a transmission time limit of: 3, 10, and 30 minutes.

the proposed framework takes into consideration the scenario-specific bandwidth constraints, it prioritizes the transmission of lower-resolution images for initial annotation. This explains the large improvement in t_{RS}_ratio compared to the baseline approach albeit with a slight penalty in $recall\_diff$. On the other hand, given that the baseline framework does not employ a streamlining strategy, it ends up always sending the still image at the highest possible resolution. This explains its slight improvement in $recall\_diff$, however this comes with a large penalty in t_{RS}_ratio. The same observations can be made at a $data\_rate$ of 88 kbps (Figure 3b). We respectively observe for the three $t_{TRlimit}$ values of 3, 10 and 30 minutes that the $recall\_diff$ worsens to 0.35, 0.148 and 0.028 but the t_{RS}_ratio improves to 25.44, 7.27 and 2.95. The streamlining strategy of the proposed framework explains its robustness against more constraint scenarios, allowing to send the same low resolution for the initial annotation step. The only downside being the reduction of t_{HUM} due to the lower bandwidth budget, explaining the slight reduction of $recall(t_{RS})$. In contrast, the baseline framework's t_{RS} performance deteriorates given that it does not adapt to the scenario-specific constraints. Figure 3c presents the scenarios at the smallest $data\_rate$ of 22 kbps. This constraint tests the limits of the proposed framework. For the scenario with a $t_{TRlimit}$ of 3 minutes, the bandwidth constraint is too restrictive, even the lowest resolution cannot be sent, leading to a $recall(t_{RS})$ of 0. For the scenarios with a $t_{TRlimit}$ of 10 and 30 minutes, we obtain a $recall\_diff$ of 0.359 and 0.115 respectively. However, we observe staggering gain in t_{RS} of 34.4 and 11.28 respectively. These very-low bandwidth scenarios clearly show the gain in t_{RS} of the proposed streamlined framework compared to the baseline framework.

6. CONCLUSION

We present a streamlined hybrid annotation framework to enhance UAV object detection in bandwidth-restricted scenarios. The framework successfully combines automated DL-based detection with human expertise to refine annotations in uncertain areas, managing to cut the decision-making time by up to 34 times for emergency response operations. Future works will focus on optimizing each of the subroutines within the proposed framework, primarily focusing on improving both the fine-tuned DL model and the tile selection strategy for human annotation. Additionally, employing a few-shot learning strategy can improve model adaptability in emergency scenarios. Furthermore, implementing an active learning layer can benefit the model's long-term training.

7. ACKNOWLEDGMENTS

The Pléiades images were downloaded from the *Pléiades 4 Belgium* platform, through an agreement with Belgian Science Policy Office (*BELSPO*). The images were labeled by the Institut Scientifique de Service Public (*ISSeP*). Tiffanie Godelaine is supported by *MedReSyst*, funded by the Walloon Region and European Union *Wallonie 2021-2027* program.

8. REFERENCES

[1] V. Papić, P. Šolić, A. Milan, S. Gotovac, and M. Polić, "High-resolution image transmission from uav to ground station for search and rescue missions planning," *Applied Sciences*, vol. 11, no. 5, pp. 2105, 2021.

[2] N. Tuśnio and W. Wróblewski, "The efficiency of drones usage for safety and rescue operations in an open area: A case from poland," *Sustainability*, vol. 14, no. 1, pp. 327, 2021.

[3] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[4] S. Zhou, C. Deng, B. Zhao, Y. Xia, Q. Li, and Z. Chen, "Remote sensing image compression: A review," in *2015 IEEE International Conference on Multimedia Big Data*, 2015, pp. 406–410.

[5] A. Indradjad, A. S. Nasution, H. Gunawan, and A. Widipaminto, "A comparison of satellite image compression methods in the wavelet domain," *Earth and Environmental Science*, vol. 280, no. 1, 2019.

[6] M. Ehrlich, L. Davis, S. Lim, and A. Shrivastava, "Analyzing and mitigating jpeg compression defects in deep learning," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 2357–2367.

[7] F. Ghazvinian Zanjani, S. Zinger, B. Piepers, S. Mahmoudpour, and P. Schelkens, "Impact of jpeg 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images," *Journal of Medical Imaging*, vol. 6, no. 02, pp. 1, 2019.

[8] K. El Khoury, M. Fockedey, E. Brion, and B. Macq, "Improved 3d u-net robustness against jpeg 2000 compression for male pelvic organ segmentation in radiotherapy," *Journal of Medical Imaging*, vol. 8, no. 04, 2021.

[9] K. El Khoury, M. Wynen, P. Maggi, B. Bach Cuadra, and B. Macq, "Improving 3d lesion segmentation robustness against image compression in multiple sclerosis," in *2024 IEEE International Symposium on Biomedical Imaging*, 2024, p. 1.

[10] J. Zhang, J. Weng, W. Luo, J. Liu, A. Yang, J. Lin, Z. Zhang, and H. Li, "Remt: A real-time end-to-end media data transmission mechanism in uav-aided networks," *IEEE Network*, vol. 32, no. 5, pp. 118–123, 2018.

[11] A. Preethy Byju, G. Sumbul, B. Demir, and L. Bruzzone, "Remote-sensing image scene classification with deep neural networks in jpeg 2000 compressed domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3458–3472, 2021.

[12] B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia, "Half a percent of labels is enough: Efficient animal detection in uav imagery using deep cnns and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9524–9533, 2019.

[13] A. Yamani, A. Alyami, H. Luqman, B. Ghanem, and S. Giancola, "Active learning for single-stage object detection in uav images," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2024, pp. 1860–1869.

[14] B. Palmaerts, B. Beaumont, D. Graur, G. Swinnen, and E. Hallot, "Oriented aircraft object detector using scaled yolov4 on very high resolution satellite and synthetic datasets," in *2023 Joint Urban Remote Sensing Event (JURSE)*, 2023, pp. 1–4.