Appendix A

Section-by-Section Writing Guidelines

1. Abstract (150-250 words)

1) **Purpose:** The abstract should summarize the entire paper in a very concise manner. It should highlight the key problem, your contribution and the main results without delving into too much detail.

2) **Tips:**

- Keep it concise; the abstract should be a snapshot of the entire paper.

- Avoid using specific numbers or technical details unless they are essential.

3) **What to include:**

- **Research Background:** Mention the significance of labeled datasets in your chosen field, briefly touching upon their necessity for supervised learning models.

   Example: "Labeled datasets are essential for training supervised learning models, which are foundational in applications like object detection, segmentation, and classification."

- **Motivation/Problem:** State the problem or gap in existing datasets (e.g., insufficient diversity in environmental conditions,

class imbalances, etc.).

Example: "Existing datasets such as COCO and Cityscapes are limited in their diversity, leading to models underperforming in real-world scenarios."

- **Key Contribution:** Highlight the novelty of your dataset (e.g., diversity of images, improved annotation methods, expanded categories, etc.).

  Example: "This paper introduces [DatasetName], a new labeled dataset designed to address these issues, providing images from various environmental conditions and enhanced annotation protocols."

- **Evaluation and Results:** Mention how you evaluated the dataset (e.g., benchmarking against state-of-the-art models) and summarize key findings.

  Example: "We demonstrate the utility of the dataset by benchmarking object detection models, showing a 15% improvement in detection accuracy under difficult weather conditions."

- **Conclusion:** Conclude by mentioning the importance and availability of the dataset.

  Example: "The dataset is freely available for research purposes and offers a valuable resource for advancing

computer vision technologies."

2. Introduction (3-4 paragraphs, 300-400 words)

1) **Purpose:** Introduce the topic, establish the importance of labeled datasets, define the problem, and preview your contributions.

2) **Tips:**

- Make the **problem statement** clear and compelling.

- Clearly identify your **contribution** to the field.

- **Engage the reader** by linking the dataset's impact to real-world applications.

3) **What to include:**

Paragraph 1 – General Context:

- Begin by explaining the role of labeled datasets in driving progress in your chosen field (e.g., computer vision).

- Mention key applications such as object detection, image segmentation, and facial recognition.

- **Example:** *"Recent advances in computer vision are heavily dependent on labeled datasets, which provide the necessary ground truth for training supervised models."*

Paragraph 2 – Gap/Problem:

- Describe the shortcomings of existing datasets (e.g., lack of diversity, insufficient representation of certain classes).

- Provide examples such as COCO, Pascal VOC, or Cityscapes, and point out where they fall short in real-world applications.

- **Example:** *"Despite the success of datasets like COCO and Cityscapes, they are limited in their coverage of specific environmental conditions, such as nighttime or poor visibility, which are critical for autonomous driving applications."*

Paragraph 3 – Motivation and Goal:

- Clearly state why your new dataset is necessary and how it addresses the issues identified in the previous section.

- Briefly describe the characteristics of your dataset and its focus.

- Example: *"To address these gaps, we developed [DatasetName], which includes images from varied weather conditions, times of day, and urban environments to ensure a more robust evaluation of object detection models."*

Paragraph 4 – Contributions (Bullet points or in-line summary):

- Enumerate the specific contributions of your paper.

- **Example:**

  The main contributions of this paper are as follows:"

a) *A comprehensive dataset with over 10,000 images and 200,000 annotations across 12 object categories.*

b) *A robust annotation protocol that ensures high consistency and quality of labeled data.*

c) *Benchmark results evaluating multiple object detection models using this dataset, demonstrating performance improvements.*

Paragraph 5 – Outline of the Paper (optional):

- You can introduce the outline of the following sections in the last paragraph of your introduction. It's optional.

- **Example:**

  "The rest of this paper is organized as follows: Section 2 introduces the related work of this paper, Section 3 describes…"

3. Related Work (3-4 paragraphs, 350-400 words)

1) **Purpose:** Review previous work, focusing on existing datasets and annotation methods, and highlight how your dataset addresses gaps or improves upon them.

2) **Tips:**

   - Organize this section to first talk about **general datasets**, then **specialized datasets**, and finally how **your work fills the gaps**.

- Use **comparisons** to draw a clear distinction between your dataset and others.

3) What to include:

Paragraph 1 – Overview of Existing Datasets:

- Start by discussing major datasets in the field (ImageNet, COCO, OpenImages, etc.).

- Mention what tasks these datasets are commonly used for (e.g., object detection, classification).

- **Example**: *"ImageNet revolutionized image classification by providing millions of labeled images across thousands of categories. However, it focuses mainly on objects in natural settings."*

Paragraph 2 – Datasets for Specific Domains:

- Provide examples of domain-specific datasets (e.g., Cityscapes for urban scene understanding, KITTI for autonomous driving).

- Briefly analyze the strengths and weaknesses of these datasets.

- **Example**: *"Cityscapes excels in urban scene understanding but lacks variety in weather conditions, limiting its applicability to real-world autonomous driving scenarios."*

Paragraph 3 – Annotation Tools and Methods:

- Discuss the various tools used for dataset annotation (e.g., CVAT, LabelMe, VIA).

- Mention how these tools impact annotation quality and consistency.

- **Example**: *"Annotations are typically generated through manual labeling using tools like CVAT, but inconsistencies arise due to inter-annotator variability, requiring quality control measures."*

Paragraph 4 – Positioning Your Work:

- Discuss how your dataset improves upon these existing efforts, addressing specific gaps such as environmental variability, object class imbalance, or labeling accuracy.

- **Example**: *"In contrast, our dataset ensures high diversity by capturing images under varying weather conditions and includes previously underrepresented object categories."*

4. **Dataset Construction (4–5 paragraphs, 450–500 words)**

1) **Purpose:** Provide a detailed description of how the dataset was created, including data collection, preprocessing, annotation, and quality assurance.

2) **Tips:**

- This section requires **specific details** about the

process, so the reader should be able to recreate it.

- Include visuals (if applicable) such as examples of annotations or images from the dataset.

- Be transparent about the **challenges faced** during construction (e.g., inter-annotator disagreements).

3) **What to include:**

Paragraph 1 – Data Collection:

- Describe the sources of data (e.g., cameras, drones, online image sources) and the variety of conditions (weather, time of day).

- **Example:** *"Images were collected using high-resolution cameras mounted on vehicles across five cities, capturing scenes during day and night, in clear weather and rain."*

Paragraph 2 – Preprocessing:

- Explain any preprocessing steps (e.g., resolution normalization, image filtering, data augmentation).

- **Example:** *"To standardize image input, all images were resized to 1280x720 pixels. Low-quality images (e.g., blurry or underexposed) were excluded based on sharpness thresholds."*

Paragraph 3 – Annotation Protocol:

- Describe the annotation process (tools used, number of

annotators, annotation types).

- **Example:** *"Each image was annotated using CVAT by a team of trained experts, who labeled objects with bounding boxes and class labels. Each object was also assigned a confidence score to indicate the certainty of the annotation."*

Paragraph 4 – Quality Control and Validation:

- Describe the steps taken to ensure high-quality annotations, including inter-annotator agreement, redundancy checks, and audits.

- **Example:** *"To ensure consistency, each image was annotated by at least two annotators, with discrepancies resolved by a third expert. Annotations were also reviewed for overlap and mislabeling."*

Paragraph 5 – Final Dataset Features:

- Summarize key characteristics of the dataset (number of images, classes, size, etc.).

- **Example:** *"The final dataset consists of 12,000 images with over 230,000 bounding box annotations across 12 object categories, with a strong emphasis on urban environments."*

5. Dataset Statistics (2–3 paragraphs, 250–300 words)

1) Purpose: Provide a statistical summary of your dataset, showing its diversity, size, and balance.

2) Tips:

- Use **quantitative data** to showcase the diversity and balance in your dataset.

- Visual aids (like **bar charts** or **pie charts**) can be useful to summarize class distributions.

- Explain **imbalances** or challenges with the dataset, and how they might affect model performance.

3)What to include:

Paragraph 1 – General Dataset Stats:

- Total number of images, annotations, and classes.

- **Example**: *"The dataset contains 12,000 images with 230,000 object annotations across 12 categories, including vehicles, pedestrians, and traffic signals."*

Paragraph 2 – Class Distribution:

- Describe how the classes are distributed and whether the dataset is balanced or imbalanced.

- **Example**: *"The dataset includes both common objects (e.g., cars, people) and less frequent categories (e.g., buses, bicycles), with a class imbalance observed in the number of instances per category."*

Paragraph 3 – Environmental Variability:

- Highlight the variability of the dataset (different weather,

lighting conditions, locations).

- **Example**: *"Images cover a wide range of lighting conditions, from bright daylight to nighttime scenes, and various weather scenarios, such as rain and fog."*

6. Evaluation (3–4 paragraphs, 350–400 words)

1) **Purpose:** Describe how you evaluated your dataset, including model performance and metrics used.

2) **Tips:**

- **Provide concrete results**: Include key numbers (e.g., mAP scores) that demonstrate the utility of your dataset.
- Discuss not only the **successful models** but also the areas where models may have **struggled**, especially with complex conditions.

3) **What to include:**

Paragraph 1 – Evaluation Setup:

- Describe the models used for evaluation (e.g., YOLOv5, Faster R-CNN) and the training/test split.
- **Example**: *"We evaluated the dataset using three state-of-the-art object detection models: YOLOv5, Faster R-CNN, and SSD. A 70/30 training-test split was applied."*

Paragraph 2 – Metrics:

- Mention the metrics used for evaluation (e.g., mAP, IoU).

- **Example**: *"Performance was evaluated using mean Average Precision (mAP) at IoU thresholds of 0.5 and 0.5:0.95."*

Paragraph 3 – Results:

- Present the key results (e.g., mAP, precision, recall).
- **Example**: *"YOLOv5 achieved an mAP@0.5 of 78%, outperforming Faster R-CNN (75%) and SSD (72%)."*

Paragraph 4 – Analysis:

- Analyze where models performed well or struggled.
- **Example**: *"Models performed well on common classes like 'car' and 'person', but struggled with small or occluded objects, especially under low-light conditions."*

7. Conclusion (1–2 paragraphs, 200–250 words)

1) Purpose: Conclude the paper, summarize key points, and discuss future work.

2) Tips:

- Keep the **conclusion concise** and focused on future opportunities.
- Avoid introducing **new information** that wasn't discussed in the main body of the paper.

3) What to include:

Paragraph 1 – Recap:

- Summarize the key points of your work.

- **Example**: *"This paper presented [DatasetName], a new labeled dataset for object detection, which addresses several limitations in existing datasets."*

Paragraph 2 – Contributions & Future Work:

- Highlight your main contributions and suggest next steps.

- **Example**: *"We believe our dataset provides a valuable benchmark for future research. In future work, we plan to expand the dataset to include semantic segmentation labels and further improve annotation quality."*