Secure Visual Data Processing via Federated Learning

Pedro Santos[1][a], Tânia Carvalho[1,2][b], Filipe Magalhães[2] and Luís Antunes[1,2][c]

[1] *Departamento de Ciência de Computadores, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, s/n, 4169– 007, Porto, Portugal*

[2] *TekPrivacy, Lda, R. Alfredo Allen n.º 455 461, 4200-135, Porto, Portugal*

Keywords: Federated Learning, Object Detection, Image Labeling, Anonymization, Data Privacy.

Abstract: As the demand for privacy in visual data management grows, safeguarding sensitive information has become a critical challenge. This paper addresses the need for privacy-preserving solutions in large-scale visual data processing by leveraging federated learning. Although there have been developments in this field, previous research has mainly focused on integrating object detection with anonymization or federated learning. However, these pairs often fail to address complex privacy concerns. On the one hand, object detection with anonymization alone can be vulnerable to reverse techniques. On the other hand, federated learning may not provide sufficient privacy guarantees. Therefore, we propose a new approach that combines object detection, federated learning and anonymization. Combining these three components aims to offer a robust privacy protection strategy by addressing different vulnerabilities in visual data. Our solution is evaluated against traditional centralized models, showing that while there is a slight trade-off in accuracy, the privacy benefits are substantial, making it well-suited for privacy sensitive applications.

1 INTRODUCTION

The rapid growth of visual data has raised significant privacy concerns, particularly in domains like healthcare, surveillance, social media, and autonomous vehicles. Regulatory frameworks such as the General Data Protection Regulation (GDPR) mandate strict protection of identifiable information, enforcing privacy standards in data handling.

Conventional approaches for managing visual information rely on centralized machine learning, where data is collected and processed in a single location. However, this setup can expose sensitive information to risks such as unauthorized access (Shokri and Shmatikov, 2015). Even anonymization techniques can be insufficient against sophisticated attacks, including model inversion and adversarial attacks, which can reconstruct sensitive information despite masking or encryption efforts (Fredrikson et al., 2015; Goodfellow et al., 2014). These challenges highlight the need for decentralized solutions that prioritize privacy and scalability (Bharati et al., 2022).

Federated Learning (FL) (McMahan et al., 2023) offers a promising alternative by enabling decentralized model training without sharing raw data, inherently enhancing privacy. However, integrating FL with AI-driven visual data tools, particularly for labeling and anonymization, remains underexplored. Most studies address privacy through anonymization (Andrade, 2024) or federated learning (Yu and Liu, 2019), but fail to combine these approaches. This leaves vulnerabilities when data must be shared or processed across multiple devices or organizations.

To bridge this gap, we propose a framework combining FL, object detection, and anonymization. Object detection identifies sensitive regions, such as faces and license plates, while FL eliminates the need to share raw data. Detected sensitive regions are then masked using anonymization techniques, providing a multi-layered defense against privacy breaches. To our knowledge, this paper presents the first approach combining these three components. Our key contributions are highlighted as follows.

- **Visual Data Labeling in a Cutting Edge FL Framework:** we use a recent and straightforward FL technology with a well-known object detection algorithm for an efficient and secure visual data labeling system.
- **Anonymization Layer:** we apply obfuscation

[a] https://orcid.org/0009-0002-8601-7905
[b] https://orcid.org/0000-0002-7700-1955
[c] https://orcid.org/0000-0002-9988-594X

Santos, P., Carvalho, T., Magalhães, F. and Antunes, L.
Secure Visual Data Processing via Federated Learning.
DOI: 10.5220/0013183000003899
In *Proceedings of the 11th International Conference on Information Systems Security and Privacy (ICISSP 2025) - Volume 2*, pages 534-541
ISBN: 978-989-758-735-1; ISSN: 2184-4356

techniques to anonymize sensitive visual data in the federated environment.

- **Performance Evaluation:** we conduct comprehensive experiments to evaluate the performance and scalability of the proposed solution.

2 LITERATURE REVIEW

This paper explores three main domains: object detection, anonymization techniques, and federated learning (FL). In this section, we review each area and highlight our contributions to the state-of-the-art.

2.1 Object Detection

Object detection is widely studied, particularly focusing on features such as color, texture, and spatial relationships to improve visual data labeling tasks (Veltkamp and Tanase, 2000; Ericsson et al., 2022; Bouchakwa et al., 2020). Recent advances emphasize self-supervised learning, which scales well with unlabeled data (Ericsson et al., 2022). Despite improvements, achieving high labeling accuracy and efficiency remains challenging (Zou et al., 2023). While automated methods dominate, manual and semi-automated labeling continue to play a role in ensuring quality, particularly in critical applications requiring high accuracy (Girshick et al., 2014). Deep learning approaches dominate modern object detection and can be classified into two categories: one-stage and two-stage algorithms (Zou et al., 2023).

One-stage detectors predict object locations and classes in a single step, optimizing speed and computational efficiency. These models are ideal for real-time tasks such as autonomous driving and surveillance. Examples include YOLO (Jocher et al., 2020), SSD (Liu et al., 2016), and RetinaNet (Lin et al., 2017). These methods leverage advanced techniques like focal loss and transformers to improve performance in diverse applications.

Two-stage detectors identify regions of interest first and refine predictions in the second step, prioritizing accuracy over speed. Examples include R-CNN (Girshick et al., 2014), Faster R-CNN (Ren et al., 2015) and Mask R-CNN (He et al., 2017). These algorithms excel in scenarios such as medical imaging and satellite imagery (Chen et al., 2019).

2.2 Anonymization Techniques

Anonymization has gained importance with stricter privacy regulations. Traditional techniques, such as blurring and masking, balance privacy and utility but can be vulnerable to attacks like re-identification (Senior, 2009; Ren et al., 2018). Advanced approaches use Generative Adversarial Networks (GANs) to replace sensitive features with synthetic versions, enhancing privacy without compromising context (Todt et al., 2022). Recent studies show that realistic methods outperform traditional anonymization in maintaining data utility (Hukkelås and Lindseth, 2023).

However, advanced methods often require higher computational resources, limiting their applicability in real-time systems. The trade-off between privacy and usability highlights the need for lightweight solutions that integrate seamlessly with decentralized learning frameworks.

2.3 Federated Learning

FL enables decentralized training of machine learning models by exchanging updates instead of raw data. This reduces privacy risks while supporting collaboration among data owners (Guan et al., 2024). The two most common FL settings include Horizontal (HFL) and Vertical Federated Learning (VFL).

HFL involves datasets with similar features but different users. For example, several hospitals, each with their patient records with similar attributes, such as blood test results or medical images, can use HFL to collaboratively train a shared model. Each hospital trains a local model on its dataset and only transmits the learned model parameters to a central server, ensuring that sensitive data is not shared (Kaissis et al., 2020). On the other hand, VFL is used when datasets share common users but different features, e.g., combining demographic and genetic data. Techniques like Private Set Intersection (PSI) ensure secure matching without revealing sensitive data (Angelou et al., 2020). Recent advancements incorporate encryption and differential privacy to strengthen security in VFL (Yang et al., 2023; Liu et al., 2024).

2.4 Current Research Directions

Efforts to integrate object detection with either FL or anonymization show promising results, but limitations remain. For instance, Andrade (2024) focus on anonymization without addressing FL, while Yu and Liu (2019) leverage FL but fail to include anonymization layers, exposing sensitive data. Similarly, Memia (2023) use YOLOv8 in FL without an anonymization step. Our work builds on these studies, especially the latter, by integrating all three components: object detection, FL, and anonymization.

To our knowledge, we are the first to combine these three components in which besides leveraging an AI model through FL, we apply an anonymization layer on the detected sensitive visual data.

3 VISUAL DATA PROTECTION AND LABELING

This section outlines the experimental evaluation procedure. We detail the methodology, providing an overview of the high-level architecture of the proposed system, including the algorithm used to train the model and a description of the dataset.

3.1 Methodology

We aim to develop a robust framework that ensures data privacy, facilitates accurate object detection, and maintains the integrity of the anonymized data for analysis. As such, we present our proposed methodology in Figure 1 [1].

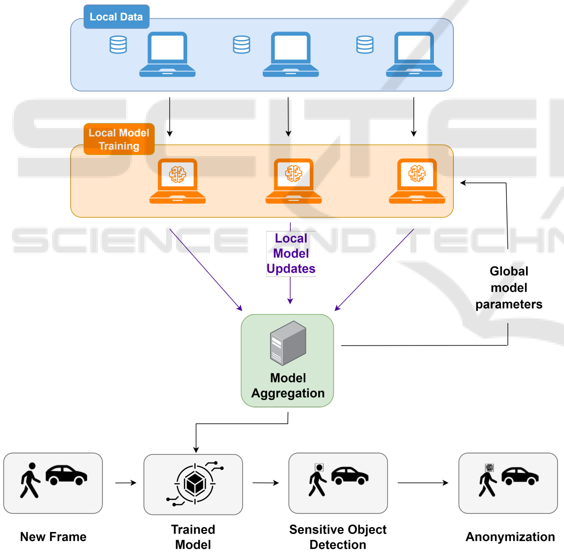

Figure 1: Methodology for sensitive data detection and anonymization using federated learning.

The overall process is divided into several steps:

- **Local Model Training:** in the first step, each participant trains a local object detection model on their private visual data.
- **Local Model Updates Transmission:** participants send only the model updates (e.g. learned weights or gradient) to a central server after training.
- **Model Aggregation:** the central server aggregates the updates from all participants to refine a global model. Federated averaging is often used to combine local models iteratively, improving the global model's performance at each training round.
- **Deployment of the Trained Model:** once the global model reaches satisfactory performance, it can be deployed for tasks like processing new images or video frames in real time, benefiting from the collective knowledge gathered during training.
- **Sensitive Data Detection:** the deployed model uses object detection algorithms to identify and label targeted classes in visual data.
- **Anonymization:** finally, after detection, the sensitive data is anonymized.

In summary, this system utilizes federated learning (FL) to train the model across multiple devices without centralizing the data. The object detection component identifies and locates sensitive areas within the visual content, such as faces or license plates. To ensure that privacy is maintained, an anonymization process will then be applied to these sensitive areas, masking or altering them to prevent identification. We aim to create a scalable and secure solution that can be used in various applications, from surveillance and healthcare to social media and content-sharing platforms, where privacy and data protection are fundamental.

3.2 Methods

Our methodology integrates FL with Flower (Beutel et al., 2020), object detection using YOLOv8 (Jocher et al., 2023), and Gaussian blur for anonymization.

Among the several object detection methods, YOLOv8 presents superior speed and accuracy, making it well-suited for real-time applications. Its adaptability across diverse datasets and scenarios ensures generalization, while its efficiency reduces computational demands, enabling deployment on edge devices. Also, the efficiency of YOLOv8 minimizes computational demands on edge devices, making the system more accessible in distributed environments that do not require high-end hardware.

We use Flower for FL due to its flexibility and compatibility with machine learning libraries like PyTorch and TensorFlow. Its efficient communication protocols minimize data transfer overhead, while

[1] Icons: "Object Detection" by Edward Boatman https://thenounproject.com/icon/object-detection-6109597/ and "Object Detection" by Nanang A Pratama https://thenounproject.com/icon/object-detection-6943616/ from Noun Project.

its modular architecture simplifies integration with YOLOv8. Flower supports dynamic participant selection, ensuring scalability and stability during training.

Two aggregation methods were tested:

- **FedAvg (McMahan et al., 2017)** which computes a simple average of updates, offering computational efficiency but slower convergence in heterogeneous settings.
- **FedOpt (Reddi et al., 2020)** which dynamically adjusts learning rates during aggregation, improving convergence and performance stability.

For anonymization, Gaussian blur was chosen for its simplicity, computational efficiency, and compatibility with FL environments. While advanced techniques like GANs or differential privacy provide stronger guarantees, their higher computational costs make them less practical for distributed systemst (Chung et al., 2024).

3.3 Data

We used the Open Images Dataset V6 Krasin et al. (2017) [2], focusing on ”Vehicle registration plate” and ”Human face” annotations. Bounding box coordinates were normalized for YOLOv8 compatibility.

Our dataset contains 29,690 images, divided into:

- **Training Set:** 14,485 images (48.8%)
- **Validation Set:** 3,848 images (13.0%)
- **Testing Set:** 11,357 images (38.2%)

To simulate a realistic FL environment, the dataset was partitioned across three participants, each accessing a distinct subset. This distribution maintains privacy and prevents centralization while reflecting real-world data partition. Thus, we can also study performance under limited data availability per participant.

4 RESULTS

This section presents the experimental evaluation of our proposed methodology. First, we establish a baseline using a centralized model to compare performance against federated learning (FL). Then, we analyze the effects of training rounds, aggregation methods, and computational costs on FL performance. Finally, we assess the anonymization layer's effectiveness in preserving privacy while retaining utility for object detection tasks.

The experiments were conducted on a high-performance virtual machine with a 16-core CPU, 128 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU (24 GB VRAM).

4.1 Baseline Model

First, we trained the YOLOv8 model on the entire dataset using a centralized approach, where all data is stored and processed in one location. Standard hyperparameters were used to optimize bounding box regression, class prediction, and distribution focal loss.

We evaluate the model's performance using precision, recall, and mean average precision (mAP). Precision measures the ratio of true positives to predicted positives, while recall captures the ratio of true positives to actual positives. The mAP evaluates performance across multiple confidence thresholds, including mAP50 and mAP50-95.

Table 1 presents the results regarding the test set. The model's performance improves as the number of epochs increases, reaching its maximum at 200 epochs. This trend indicates that the model was successfully learning and optimizing over time. However, we observe that using 100 epochs, the centralized model provides lower loss, but the remaining metrics are generally slightly worse than training the model with 150 epochs.

4.2 Federated Learning Setting

We now evaluate the performance of the proposed FL framework by analyzing key factors such as training epochs, aggregation methods, and communication costs in comparison with the baseline.

4.2.1 Epoch Variation on Model Performance

The number of epochs represents a critical hyperparameter in machine learning, as it determines the number of times the learning algorithm will work through the entire training dataset. Typically, increasing the number of epochs can lead to enhanced model performance, as it allows the model more opportunities to adjust its parameters and reduce errors. The baseline results confirm this. However, in FL, increasing the number of epochs may result in minimal performance gains while considerably extending training time due to communications in this procedure. Table 2 shows the results using the same epoch variation as previously and the aggregation method FedAvg.

In general, we verify the same trend as before: the higher the number of epochs, the better the performance of the federated model. This is particularly ev-

[2] Publicly accessible at https://storage.googleapis.com/openimages/web/visualizer/index.html

Table 1: Performance metrics for the baseline model (centralized) across varying epochs.

Epochs	mAP50	mAP50-95	Recall	Precision	Loss	Training Time (sec)
25	76.07%	51.21%	72.21%	81.00%	0.00039	1,740
50	78.52%	54.01%	75.46%	83.20%	0.00039	3,357
100	79.41%	55.25%	76.44%	82.01%	0.00036	6,421
150	79.69%	55.36%	76.43%	84.21%	0.00038	9,910
200	80.05%	56.11%	77.34%	84.32%	0.00037	12,760

Table 2: Performance metrics results for the federated model with 5 rounds across varying epochs.

Epochs	mAP50	mAP50-95	Recall	Precision	Loss	Training Time (sec)
25	62.55%	43.79%	48.71%	75.03%	0.00040	8,047
50	56.23%	38.69%	51.09%	67.53%	0.00042	15,606
100	63.77%	44.85%	50.84%	78.85%	0.00040	30,359
150	64.70%	45.01%	58.88%	72.12%	0.00040	38,183
200	74.68%	52.32%	68.52%	77.82%	0.00042	49,633

ident in mAP50, mAP50-95 and recall, which compared to the baseline, these metrics present higher differences between the 25 and 200 epochs. Nevertheless, we observe some fluctuations, especially in precision with an increase of less than 3% between 25 epochs and 200 epochs. Also, the metric loss presents higher fluctuations. These outcomes may be attributed to the communication and aggregation model in FL. Despite the reached peak at 200 epochs, the federated model generally underperforms the baseline. This may be attributed to the distributed nature of data in FL, which may hinder its ability to generalize as effectively as the baseline model.

Consequently, the run time for 200 epochs is four times longer than the baseline. Therefore, it is essential to assess the trade-off between the utility gains over higher epochs and the computational resources required to achieve them, making it crucial to tailor the choice of epochs to the specific needs and constraints of the targeted application.

4.2.2 After-Effect of Adding More Rounds

We also analyzed the impact of increasing communication rounds, setting the number of rounds to 3, 4, 5, 6, 7, and 8, with 200 training epochs. This setup evaluates how performance evolves with successive rounds of communication between edge devices and the central server. Table 3 summarizes the results.

Performance generally improves with more rounds, particularly between 3 and 5 rounds, where mAP50 and recall increase by over 40%. However, after 5 rounds, gains diminish, with mAP50 and mAP50-95 rising by less than 1.5% and precision improving only marginally (0.37%).

Training time also increases with more rounds. Between 3 and 5 rounds, runtime grows by approximately 16.8 seconds, while adding rounds up to 8 increases runtime by another 18 seconds. Despite nearing baseline performance, the computational costs for 8 rounds are nearly 5 times higher than centralized training. These results highlight a trade-off between performance improvements and resource demands, emphasizing the need to balance communication rounds with computational efficiency.

4.2.3 Aggregation Methods

Focusing on the best results achieved previously, we compare the performance of FedAvg and FedOpt. Table 4 presents the comparative analysis of the performance of such aggregation methods over 8 rounds.

In general, both methods demonstrate similar performance values. The maximum is reached at 200 epochs. However, FedOpt shows a consistent improvement of the loss metric over epochs, while FedAvg shows the same behavior for mAP50. Additionally, we observe that FedOpt presents a better mAP50-95, precision and loss. In contrast, FedAvg presents better results for the remaining metrics.

Despite the high similarity between the two aggregation methods, we continue to experiment with FedOpt due to its superior optimization techniques, especially when dealing with potential variations in data quality between participants. FedOpt allows for more efficient convergence and provides stability in training. These characteristics make it more suitable for scenarios where the quality and availability of participant data may be uncertain.

4.3 Anonymization Layer

After the best-federated setting selection, we demonstrate the effectiveness of the anonymization layer. Thus, we obfuscated detected sensitive regions, such as faces and license plates, using Gaussian blur. Figure 2 illustrates the process, highlighting anonymized areas with green bounding boxes.

Table 3: Performance metrics results for the federated model with 200 epochs over different rounds.

Round	mAP50	mAP50-95	Recall	Precision	Loss	Training Time (sec)
3	28.84%	19.21%	24.30%	36.17%	0.00040	32,834
4	56.41%	40.95%	50.73%	76.77%	0.00039	43,386
5	74.68%	52.32%	68.52%	77.82%	0.00042	49,633
6	74.42%	52.29%	68.09%	78.76%	0.00040	60,545
7	76.01%	53.44%	71.09%	77.75%	0.00039	61,601
8	76.69%	53.57%	71.27%	73.92%	0.00039	67,798

Table 4: Performance metrics results using FedAvg and FedOpt over different epochs and 8 rounds.

Method	Epochs	mAP50	mAP50-95	Recall	Precision	Loss	Training Time (sec)
FedOpt	25	73.35%	52.93%	70.32%	84.36%	0.00043	13,328
	50	72.34%	50.85%	66.54%	76.01%	0.00042	25,199
	100	75.11%	52.42%	66.05%	75.69%	0.00041	48,591
	150	73.62%	51.49%	67.86%	76.47%	0.00040	57,734
	200	76.31%	53.81%	70.83%	79.14%	0.00037	68,205
FedAvg	25	71.27%	52.23%	67.34%	76.17%	0.00041	13,307
	50	72.48%	50.93%	68.49%	74.64%	0.00040	25,176
	100	75.62%	53.47%	69.58%	80.26%	0.00042	48,317
	150	76.51%	53.24%	69.43%	73.75%	0.00040	58,780
	200	76.69%	53.57%	71.27%	73.92%	0.00039	67,798

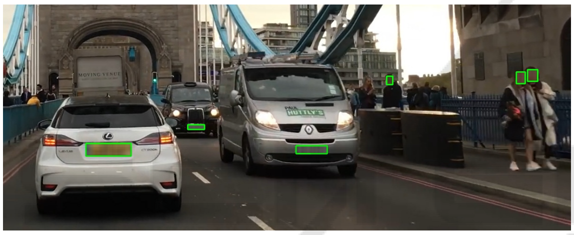

Figure 2: A visual representation of the anonymization layer applied to an image. The green bounding boxes highlight the areas where anonymization was applied, specifically targeting license plates and faces.

While the anonymization layer protects privacy, it is also essential to maintain the data utility for downstream object detection. The anonymization process was carefully designed to minimize the loss of critical information that is necessary for model performance. For example, blurring license plates retains the overall structure of vehicles, allowing the model to maintain detection accuracy for scene analysis.

In summary, integrating object detection with FL enables secure visual data processing without transferring raw data. While some performance loss is observed compared to centralized models, these losses are minor. The anonymization layer further strengthens privacy protection by masking sensitive features effectively, as shown in Figure 2.

4.4 Discussion

This section discusses the challenges and implications of deploying federated learning (FL) systems, focusing on scalability, communication efficiency, ethical considerations, and privacy risks. We also highlight limitations and propose directions for future research.

Performance Loss and Detection Risks. FL introduces slight performance losses compared to centralized models. However, our results show that the loss is marginal. For example, the mAP50 of the federated model (76.69%) is close to the baseline (80.05%). Optimizing hyperparameters and aggregation methods mitigates risks of missed detections. Qualitative evaluation confirms that the anonymization layer effectively masks sensitive features while preserving contextual information for downstream tasks.

Scaling Participants. Increasing the number of participants can improve performance by using more data, but it also introduces challenges. For research purposes, data was divided among participants, reducing dataset size per node and slightly lowering accuracy compared to centralized training. In real-world applications, data scarcity can be mitigated using techniques such as data augmentation or federated data synthesis to enhance model performance.

Communication Overhead. FL requires participants to exchange model updates, which may lead to increased communication costs as the number of rounds or participants grows. Table 3 highlights this compromise between performance and computational resources. Future work should investigate techniques such as model compression and asynchronous updates (Wang et al., 2024) to reduce overhead, particularly in resource-constrained environments.

Scalability and Deployment Challenges. Scaling FL systems raises issues such as participant dropout, variable participation rates, and inconsistent data availability (Memia, 2023). These factors can affect model performance and integrity. Future work should explore fault-tolerant algorithms and dynamic participant management strategies to handle such variability. Regulatory compliance, particularly with frameworks like GDPR, must also be addressed to ensure deployments align with legal standards while maintaining privacy protections.

Ethical Considerations and Bias. FL models may amplify biases present in training data (Mohri et al., 2019), leading to unfair outcomes. This is high relevant in the context of visual data, where biases can result in the misidentification or inadequate anonymization of specific ethnic groups, potentially leading to unfair outcomes. Future research should focus on developing bias mitigation techniques and fairness-aware learning methods to prevent disparities. Also, transparency in model development, including data sources and bias mitigation measures, is critical to building trust and ensuring ethical outcomes.

Privacy Risks. Although anonymization masks identifiable features, risks of re-identification remain, especially if unique visual markers, such as clothing patterns or tattoos, are not sufficiently obfuscated (Fredrikson et al., 2015). Advanced anonymization techniques like synthetic data generation and differential privacy should be explored to address these vulnerabilities (Hukkelås and Lindseth, 2023). Future research could also focus on AI models capable of detecting and masking unique visual characteristics automatically, further strengthening privacy protections.

5 CONCLUSIONS

This paper presents the first framework that integrates object detection and federated learning with anonymization to enhance privacy in visual data management. By enabling decentralized model training, FL ensures privacy without sharing raw data, while YOLOv8 achieves efficient object detection of sensitive information. Our results demonstrate that despite some performance losses of the federated model compared to the baseline model (centralized), our solution is highly efficient in detecting critical regions such as faces and license plates. The detected regions are protected with an anonymization layer that effectively masks these identifiable features.

ACKNOWLEDGMENTS

The work of Luís Antunes is supported through the Operational Competitiveness and Internationalization Programme (COMPETE 2030) [Project Atlas].

REFERENCES

Andrade, R. (2024). Privacy-preserving face detection: A comprehensive analysis of face anonymization techniques. Master's thesis, Universidade do Porto (Portugal).

Angelou, N., Benaissa, A., Cebere, B., Clark, W., Hall, A. J., Hoeh, M. A., Liu, D., Papadopoulos, P., Roehm, R., Sandmann, R., Schoppmann, P., and Titcombe, T. (2020). Asymmetric private set intersection with applications to contact tracing and private vertical federated machine learning.

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Kwing, H. L., Parcollet, T., Gusmão, P. P. d., and Lane, N. D. (2020). Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*.

Bharati, S., Mondal, M., Podder, P., and Prasath, V. (2022). Federated learning: Applications, challenges and future directions. *International Journal of Hybrid Intelligent Systems*, 18(1-2):19–35.

Bouchakwa, M., Ayadi, Y., and Amous, I. (2020). A review on visual content-based and users' tags-based image annotation: methods and techniques. *Multimedia Tools and Applications*.

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W., et al. (2019). Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4974–4983.

Chung, J., Hyun, S., Shim, S.-H., and Heo, J.-P. (2024). Diversity-aware channel pruning for stylegan compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7902–7911.

Ericsson, L., Gouk, H., Loy, C. C., and Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, 39(3):42–62.

Fredrikson, M., Jha, S., and Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pages 1322–1333.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Guan, H., Yap, P.-T., Bozoki, A., and Liu, M. (2024). Federated learning for medical image analysis: A survey.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

Hukkelås, H. and Lindseth, F. (2023). Does image anonymization impact computer vision training? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 140–150.

Hukkelås, H. and Lindseth, F. (2023). Does image anonymization impact computer vision training? *arXiv preprint*.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics YOLO. Available at https://github.com/ultralytics/ultralytics.

Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Diaconu, L., Ingham, F., Poznanski, J., Fang, J., Yu, L., et al. (2020). ultralytics/yolov5: v3. 1-bug fixes and performance improvements. *Zenodo*.

Kaissis, G. A., Makowski, M. R., Rückert, D., and Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6):305–311.

Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Veit, A., et al. (2017). Openimages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://github. com/openimages*, 2(3):18.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer.

Liu, Y., Kang, Y., Zou, T., Pu, Y., He, Y., Ye, X., Ouyang, Y., Zhang, Y.-Q., and Yang, Q. (2024). Vertical federated learning: Concepts, advances, and challenges. *IEEE Transactions on Knowledge and Data Engineering*.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2023). Communication-efficient learning of deep networks from decentralized data.

Memia, A. (2023). Federated learning for edge computing: Real-time object detection. Master's thesis, University of Skövde, Skövde, Sweden. http://hh.divaportal.org/smash/get/diva2:1785124/FULLTEXT01.pdf.

Mohri, M., Sivek, G., and Suresh, A. T. (2019). Agnostic federated learning.

Reddi, S., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečnỳ, J., Kumar, S., and McMahan, H. B. (2020). Adaptive federated optimization. *arXiv preprint arXiv:2003.00295*.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Ren, Z., Lee, Y. J., and Ryoo, M. S. (2018). Learning to anonymize faces for privacy preserving action detection. In *Proceedings of the european conference on computer vision (ECCV)*, pages 620–636.

Senior, A. (2009). *Protecting privacy in video surveillance*. Springer.

Shokri, R. and Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, page 1310–1321, New York, NY, USA. Association for Computing Machinery.

Todt, J., Hanisch, S., and Strufe, T. (2022). Fantômas: Understanding face anonymization reversibility. *arXiv preprint arXiv:2210.10651*.

Veltkamp, R. and Tanase, M. (2000). Content-based image retrieval systems: A survey. *researchgate.net*.

Wang, L., Zhou, H., Bao, Y., Yan, X., Shen, G., and Kong, X. (2024). Horizontal federated recommender system: A survey. *ACM Computing Surveys*, 56(9):1–42.

Yang, L., Chai, D., Zhang, J., Jin, Y., Wang, L., Liu, H., Tian, H., Xu, Q., and Chen, K. (2023). A survey on vertical federated learning: From a layered perspective. *arXiv preprint arXiv:2304.01829*.

Yu, P. and Liu, Y. (2019). Federated object detection: Optimizing object detection model with federated learning. In *Proceedings of the 3rd international conference on vision, image and signal processing*, pages 1–6.

Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276.