Machine Learning with Applications 23 (2026) 100806

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

AdAPT: Advertisement detector adaptation under newspaper domain shift with null-based pseudo-labeling

Faeze Zakaryapour Sayyad [a,b,*], Tobias Pettersson [c,d,e], Seyed Jalaleddin Mousavirad [b], Irida Shallari [b], Mattias O'Nils [b]

[a] Media Research AB, Ringvägen 31, Östersund, 831 37, Sweden
[b] Mid Sweden University, Holmgatan, Sundsvall, 852 30, Sweden
[c] ITAB Shop Products AB, Instrumentvägen 2, Jönköping, 553 02, Sweden
[d] University of Skövde, Högskolevägen 1, Skövde, 541 28, Sweden
[e] Jönköping University, Gjuterigatan 5, Jönköping, 553 18, Sweden

ARTICLE INFO

Keywords:
Cross-domain advertisement detection
Deep learning
Object detection
Domain adaptation
Pseudo labeling

ABSTRACT

Detecting advertisements in digitized newspapers is a key step in large-scale media analytics and digital archiving. However, variations in layout, typography, and advertisement design across publishers and time periods cause significant domain shifts that reduce the generalization ability of supervised detectors. This paper presents AdAPT, a confidence-guided pseudo-labeling pipeline for unsupervised domain adaptation in advertisement detection. The proposed method leverages both advertisement-free (Null) and advertisement-containing pages from unlabeled target domains to generate reliable pseudo-labels. By retraining a YOLO-based detector using labeled source data combined with filtered pseudo-labeled target samples, AdAPT achieves robust adaptation without requiring manual annotation. Experiments conducted on two unseen newspapers (Adresseavisen and iTromsø) demonstrate that Null-based pseudo-labeling provides the most stable and accurate adaptation, yielding up to 38% error reduction compared to the baseline. The results highlight AdAPT as a simple, scalable, and annotation-efficient solution for maintaining high-performance advertisement detection across diverse newspaper collections.

1. Introduction

The automated detection of advertisements (Ads) in printed newspapers is a practical yet challenging task in document image analysis. From a computer vision perspective, Ads are visually diverse objects that vary in size, layout, typography, and graphic design, making them harder to detect reliably than structured elements such as text blocks or tables. Accurate detection is critical because it enables downstream applications including large-scale media analysis, content categorization, and revenue modeling in digital archives. However, acquiring labeled data for every new newspaper collection is costly and labor-intensive, restricting the scalability of fully supervised detection pipelines. A more robust detector that can generalize across heterogeneous newspaper styles would not only reduce annotation and retraining costs but also address broader challenges in cross-domain object detection such as domain adaptation, class imbalance, and reliability under distribution shifts that remain open in computer vision.

Detecting Ads in newspapers is challenging because they often resemble editorial content. Their placement is unpredictable appearing at the bottom, embedded within article columns, or spanning an entire page and modern Ads increasingly mimic the visual style of articles, blurring editorial and commercial boundaries. Such diversity requires detection methods that capture both visual and contextual cues, while language differences and regional editorial practices further reduce generalizability across publishers.

To situate newspaper ad detection within the broader literature, prior work has examined ads across multiple media platforms. In television (Rondan et al., 2025; Wardana & Wibowo, 2023), radio (Álvarez et al., 2024), and digital platforms (Carvalho et al., 2021; Jain et al., 2024), detection systems often rely on structured cues such as logos, transitions, or audio patterns. For print media, related efforts have emerged from document layout analysis and segmentation (Almutairi & Almashan, 2019; Bansal et al., 2021), where Ads are usually treated as secondary elements. In parallel, methods like ADVISE (Ye & Kovashka, 2018), AdSegNet (Dhang et al., 2024), and LaBINet (Dhang et al., 2025) address Ad understanding in natural scenes, with ADVISE focusing on interpreting Ad messages and the latter two on segmentation, but all assume Ads are already localized. Similarly, PTPNet (Madi et al., 2021)

* Corresponding author at: Mid Sweden University, Holmgatan, Sundsvall, 852 30, Sweden.
E-mail addresses: Faeze.zakaryapoursayyad@miun.se (F.Z. Sayyad), tobias.pettersson@itab.com (T. Pettersson), Seyedjalaleddin.mousavirad@miun.se (S.J. Mousavirad), Irida.shallari@miun.se (I. Shallari), Mattias.onils@miun.se (M. O'Nils).

https://doi.org/10.1016/j.mlwa.2025.100806
Received 16 October 2025; Received in revised form 21 November 2025; Accepted 30 November 2025
Available online 1 December 2025

operates as a single-instance polygon regressor on Ad-present inputs rather than a full-page Ad detector. In addition, works such as Almgren et al. (2018) or the magazine 'AD or Non-AD' classifier approach the task as binary region classification (Ad vs. non-Ad). As they lack an explicit page-level localization stage, these approaches are not readily applicable to dense, layout-rich newspaper pages.

Recent efforts have applied deep object detection methods to the problem of Ad identification in digitized newspapers. Jonsson (Jonsson, 2022) employed a Faster R-CNN-based approach, showing that convolutional detectors can localize Ads with high accuracy in specific collections. More recently, AdVision (Sayyad et al., 2025) introduced a YOLOv8-based detector, demonstrating high performance across several newspaper datasets. While these methods establish the effectiveness of deep learning for Ad detection, Sayyad et al. (2025) highlights that domain shifts between newspapers and differences in layout, typography, and Ad styles — remain a barrier. Beyond cross-publisher shift, newspaper Ad distributions evolve over time due to seasonal campaigns, redesigns, and changing layout rules. In practice, detectors trained on recent data can degrade measurably over a span of months as styles shift, requiring frequent refreshes to sustain quality. Continuous fully supervised updates are impractical because they imply recurring annotation cycles. This makes low-touch adaptation strategies such as pseudo-labeling on fresh unlabeled pages attractive: they enable periodic updates that track distribution drift while minimizing labeling overhead.

Unsupervised domain adaptation (UDA) provides a promising path for overcoming both cross-publisher and temporal gaps by leveraging unlabeled target data. In general, UDA seeks to transfer knowledge from a labeled source domain (e.g., one newspaper collection) to an unlabeled target domain (e.g., a different publisher or a later time period) where annotation is unavailable. The core challenge is to mitigate the distribution shift between source and target domains so that a model trained on the source can generalize effectively to the target. A wide range of UDA strategies have been explored in computer vision, including feature alignment methods (He et al., 2023) that reduce domain discrepancy in latent space, adversarial approaches (Lv, Xia et al., 2025) that enforce domain-invariant representations, and self-training methods (Zhu et al., 2025) that iteratively refine the model using pseudo-labels on target data.

Among UDA strategies, pseudo-labeling has proven effective in related computer vision tasks by generating high-confidence predictions on unlabeled samples that are then used as additional training samples (Guo et al., 2025; Kage et al., 2024). However, its potential for Ad detection in digitized newspapers remains unexplored (Oza et al., 2023). This paper investigates how pseudo-labeling can be exploited to adapt Ad detection models across diverse newspaper domains without requiring additional annotations. To address these challenges, we propose a Confidence-guided pseudo-labeling pipeline for unsupervised domain adaptation in newspaper Ad detection.

Our main contributions are as follows:

1. We introduce AdAPT, a confidence-guided pseudo-labeling pipeline for adapting Ad detectors to new newspaper domains without requiring manual target annotations.
2. We present the first systematic study of pseudo-labeling in document Ad detection. Our results show that Null page supervision plays an important role in cross-domain generalization, outperforming Ad-only Pseudo Labeling.
3. We complement this with an analysis of limited supervised adaptation, quantifying how much labeled target data is needed to achieve performance comparable to pseudo-labeling, thereby providing practical guidance for choosing between annotation and pseudo-labeling.

Our findings demonstrate that confidence-aware self-training is a practical, effective, and easily integrable solution for scalable Ad detection across diverse newspaper formats. The remainder of the paper reviews related work (Section 2), presents our confidence-guided pseudo-labeling method (Section 4), reports results and analysis (Section 6), discusses implications and limitations (Section 7), and concludes (Section 8).

2. Related work

Advertisement detection has been explored across diverse media, each presenting distinct challenges. In broadcast domains such as television and radio, detection relies on structured cues like transitions, jingles, or logos (Álvarez et al., 2024; Rondan et al., 2025; Wardana & Wibowo, 2023), while in digital and outdoor media, visual regularities allow segmentation or insertion-based approaches (Carvalho et al., 2021; Dhang et al., 2024, 2025; Jain et al., 2024). Methods such as ADVISE (Ye & Kovashka, 2018) and PTPNet (Madi et al., 2021) focus on message interpretation or single-instance localization, assuming Ads are already isolated.

In contrast, printed newspapers present a more complex case: Ads appear within dense page layouts and often mimic editorial design. Early document layout analysis work (Almutairi, 2025; Almutairi & Almashan, 2019; Bansal et al., 2021) treated Ads as secondary elements, and image-level classifiers (Almgren et al., 2018; Jain et al., 2021) lacked spatial localization. Recent approaches using deep object detectors have advanced the field – Jonsson (Jonsson, 2022) employed Faster R-CNN, and YOLO-based methods (Sayyad et al., 2025; Zakaryapour Sayyad et al., 2024) achieved strong accuracy – but cross-domain generalization remains limited due to layout and style variation across publishers.

This issue reflects the broader challenge of domain adaptation in computer vision, where models trained on a labeled source domain may struggle to generalize to a target domain with different visual or structural characteristics. This has motivated research into adaptation strategies in computer vision, particularly unsupervised domain adaptation (UDA). UDA seeks to mitigate such gaps by exploiting unlabeled target data, reducing the need for costly annotation (Alijani et al., 2024). One strategy within UDA is pseudo-labeling, in which a source-trained model generates predictions for target samples, and reuses confident predictions as surrogate labels (Kage et al., 2024). This form of self-training has shown encouraging results in related areas such as object detection in natural images (Xu et al., 2021) and semantic segmentation (He et al., 2021), as it enables models to progressively align with the target distribution.

Recent research has extended pseudo-labeling to object detection, where bounding-box localization and class confidence must be estimated jointly. Early semi-supervised frameworks such as STAC (Sohn et al., 2020), Unbiased Teacher (Liu et al., 2021), and Soft Teacher (Xu et al., 2021) introduced teacher–student (strong/weak) consistency learning to leverage unlabeled images through high-confidence pseudo-boxes. Subsequent methods including DenseTeacher (Zhou et al., 2022), LabelMatch (Chen et al., 2022), and Consistent-Teacher (Wang et al., 2023) improve pseudo-label quality through dense prediction supervision, label matching across classification and localization, and mechanisms to reduce temporal inconsistencies in pseudo-targets. These semi-supervised detection frameworks share conceptual foundations with pseudo-label-based domain adaptation, where confident target predictions serve as auxiliary supervision to bridge domain gaps.

Transformer-based detectors have further advanced pseudo-supervision. Sparse Semi-DETR (Shehzadi et al., 2024) introduces query-refinement and reliable pseudo-label filtering to improve DETR-style detection under limited annotation, while the more recent STEP-DETR (Shehzadi et al., 2025) extends this line with a dual-teacher ensemble and pseudo-label-guided text queries to enhance class balance

and foreground–background discrimination. These methods emphasize pseudo-label reliability within transformer decoders, demonstrating that refined teacher signals and denoising strategies substantially improve detection quality.

In domain adaptation, pseudo-labeling is often combined with feature alignment. Approaches such as ConfMix (Mattolin et al., 2023), Align and Distill (Kay et al., 2024), and Dual Probabilistic Alignment (Zheng et al., 2025) jointly exploit detector predictions on the target domain and domain-invariant feature learning to mitigate distribution gaps. More challengingly, recent works address pseudo-labeling under open-set or class-imbalanced conditions, for example in domain-adaptive 3D detection with class-balanced pseudo boxes (Chen et al., 2023) and class-balanced domain adaptive object detection in driving scenes (Cai et al., 2024; Kennerley et al., 2024).

These studies establish pseudo-labeling as an effective strategy for bridging domain gaps in natural image detection. However, most rely on multi-round teacher–student pipelines that place stronger emphasis on refining classification confidence than on explicitly improving localization accuracy and they are optimized for natural-image scenes instead of structured layouts such as newspapers. In contrast, our method adopts a simpler, single-pass pseudo-labeling strategy that directly targets spatial consistency and region localization, making it more suitable for document and newspaper archives.

Beyond layout segmentation baselines, recent document-domain detectors explore semi-supervised training directly within Document Layout Analysis (DLA). PP-DocLayout (Sun et al., 2025) introduces a unified document layout detector suite (L/M/S) over 23 categories spanning papers, magazines, and newspapers, coupling a high-precision teacher (RT-DETR-L) with lightweight students via class-adaptive pseudo-labeling. The teacher provides pseudo-boxes whose thresholds are tuned per class to maximize validation F1, and students inherit these labels to trade accuracy for speed. Compared to DocLayout-YOLO (Zhao et al., 2024), PP-DocLayout emphasizes teacher–student pseudo-label transfer and finer label granularity; however, differences in taxonomies and datasets limit direct mAP comparisons. While not Ad-specific, it demonstrates that pseudo-labeling within document layout detection scales effectively across diverse page types, complementing our focus on Ad detection under domain shift.

A related line of research investigates pseudo-labeling as a means of stabilizing semi- and unsupervised training across vision tasks. In semantic segmentation, U^2PL (Wang et al., 2022) demonstrates that pixels with low confidence can still serve as negative pseudo-labels, helping to regularize decision boundaries. In facial expression recognition, SPS-NM (Lv, Ren et al., 2025) explicitly balances reliable positive pseudo-labels with multiple negatives to mitigate noise, while ClusMatch (Wu et al., 2025) shows that unified positive/negative pseudo-label learning improves clustering quality. Other works refine pseudo-labels under domain shift, for example by enforcing cross-scale and temporal consistency in object detection (Guo et al., 2025) or denoising pseudo-identities in person re-identification (Wang, Zhang et al., 2022). Although these approaches differ in task and modality, they share the insight that negatives are as important as positives when pseudo-labels are imperfect.

Similar observations arise outside the pseudo-labeling setting: in ultrasound lesion detection, Yu et al. (2023) explore negative temporal contexts to suppress false positives in real time, while Nogueira-Rodríguez et al. Nogueira-Rodríguez et al. (2023) show that incorporating explicit negative samples improves polyp detection in colonoscopy. Both works highlight that systematically modeling negatives can substantially reduce false alarms, a principle we extend to Ad detection by exploiting abundant Null pages as reliable negative supervision. Our setting departs from this convention: unlike relative negatives (e.g., "not class A") derived from ambiguous predictions, newspapers provide abundant and well-defined Null pages that contain no Ads at all. Leveraging these absolute negatives as primary pseudo-labels offers a domain-specific path to robust adaptation, complementing prior work that focused on mining negatives from within uncertain predictions.

To the best of our knowledge, pseudo-labeling has not been systematically investigated for Ad detection in documents. Existing Ad detectors such as AdVision remain purely supervised, leaving open the question of whether lightweight pseudo-labeling can extend their generalization to new newspaper datasets. Unlike natural scene datasets, newspapers combine dense text, structured layouts, and visually ambiguous elements, raising open questions about which pseudo-labels (e.g., Ad-containing pages vs. Null pages) provide the most reliable adaptation cues. Exploring pseudo-labeling in this context is therefore relevant not only for reducing annotation costs in new newspaper datasets but also for broadening the application of UDA techniques in the domain of document image analysis.

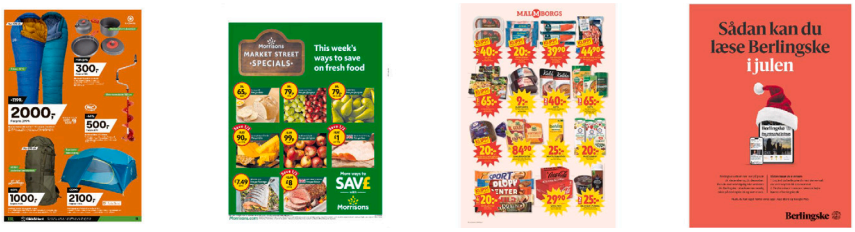
Our focus is a simple, single-pass alternative tailored to newspapers: instead of multi-round teacher–student training, we (i) generate pseudo-labels once, (ii) separate positive Ad pages from Null page, and (iii) retrain from scratch on source data mixed with pseudo-labeled target pages. This design is motivated by two properties of real archives: Null pages are plentiful and diverse, providing reliable negative supervision, whereas high-threshold Ad pseudo-boxes are sparse and still noisy under domain shift. We therefore analyze confidence thresholds for both and show that exploiting Null pages as primary pseudo-labels offers a data-efficient, compute-light path to robust cross-newspaper generalization.

3. Challenge

The main challenge in this research is achieving robust cross-domain generalization of AdVision (Sayyad et al., 2025), a YOLO-based Ad detector, across diverse digital newspaper styles. Newspaper layouts, styles, and Ad formats vary widely across publications, making it difficult for models trained on one dataset to accurately detect Ads in a different newspaper (Sayyad et al., 2025). A model trained on one publication often fails to generalize to another due to this domain shift, resulting in significant drops in detection accuracy when applied to unseen formats. Additionally, labeling new target data is costly and time-consuming, creating a bottleneck for scalable deployment. Reducing dependence on manually annotated data while ensuring reliable Ad detection is a critical challenge.

These domain shifts arise directly from the wide visual and structural diversity of newspaper Ads. Fig. 1 illustrates representative full-page Ads from Adresseavisen (Adresseavisen, 2022–2023), UK Metro (Metro, 2022–2023), Sydsvenskan (Sydsvenskan, 2022–2023), and Berlingske (Berlingske, 2022–2023). For example, Fig. 1(a) features large, distinct products with bold pricing, while Fig. 1(b) follows a neat grid layout with clear product–price pairs. In contrast, Fig. 1(c) is much denser and cluttered, with overlapping starbursts and irregular alignments. Fig. 1(d) is text-heavy with minimal product imagery, resembling a poster or branding campaign rather than a traditional product Ad. This variability illustrates why a single Ad detector may not generalize equally well across all pages, requiring adaptability to both structured and unstructured designs.

Detecting in-page Ads across newspapers is further complicated by placement and size variance as well as stylistic similarity to editorial content. As illustrated in Fig. 2, Ads can appear as right-rail blocks (e.g., UK Metro), left-rail vertical strips (e.g., Sydsvenskan), bottom-of-page strips integrated into the layout (e.g., Adresseavisen), or mid-page panels adjacent to headlines (e.g., Berlingske). These occupy different aspect ratios (column, half-page, full-height rail, bottom banner) and use diverse color schemes (from subdued editorial tones to saturated promotional palettes). Notably, the Adresseavisen example shows editorial and Ad modules sharing very similar typography and grid rules, increasing the risk of false positives when an Ad detector relies on layout cues alone. Together, variability in position, size, aspect ratio,



(a) Adresseavisen (b) UK Metro (c) Sydsvenskan (d) Berlingske

Fig. 1. Examples of full-page Ads with diverse layouts. Variability in density, typography, and visual emphasis illustrates why detectors trained on one archive may fail to generalize.

(a) Berlingske — mid-page panel right of article (b) UK Metro — right-rail Ad block (c) Sydsvenskan — full-height right rail (d) Adresseavisen — bottom strip; editorial visually similar (e) Sydsvenskan — left-rail vertical strip

Fig. 2. In-page Ads vary in placement and size: right/left rails, bottom strips, and mid-page panels, in target often with aspect ratios and styles that closely match editorial modules.

language/typography, and adjacency to editorial elements means detectors must be robust to both where an Ad lives on the page and how closely it mimics surrounding content.

Pseudo-labeling and domain adaptation techniques offer promising alternatives, but introduce new complexities, such as determining the optimal confidence threshold for pseudo-labels and assessing whether including Null (non-Ad) pages improves detection accuracy. Furthermore, selecting the best approach – whether fine-tuning an existing model, training a separate one, or using a hybrid strategy – remains an open question, especially in scenarios with limited labeled target data.

Addressing these challenges requires a systematic exploration of pseudo-label strategies, data selection criteria, and target data volume to balance accuracy and annotation cost.

4. Methodology

The main goal of this paper is to adapt AdVision, a YOLOv8-based Ad detector, to new newspaper domains without relying on extensive manual annotation. To this end, we introduce AdAPT, a single-pass pseudo-labeling pipeline in which pseudo-labels are generated once by a source-trained model and then used to retrain a detector from scratch. AdAPT is model independent; it operates at the pseudo-labeling and data-adaptation levels and can be combined with any object detection architecture without modification. AdVision is built on YOLOv8, where each input page image is processed as a whole, and the detector outputs bounding boxes, and confidence scores. In AdAPT, we use this same architecture, but we modify the training pipeline by augmenting its inputs with pseudo-labeled.

Training data therefore consists of the original labeled source set augmented with pseudo-labeled target pages. Depending on the augmentation strategy, the source labeled training set is extended with: (i) pages containing at least one Ad filtered with high-confidence pseudo-labels, (ii) Null page (pages containing no Ads), or (iii) a combination of both types. In addition to AdAPT, we also consider a limited supervised adaptation setting, where instead of pseudo-labels we augment the source dataset with a small amount of manually annotated target-domain data. This variant serves as a baseline for quantifying how much labeled target data is required to achieve adaptation when introducing a new newspaper.

We formalize the datasets in terms of a labeled source domain and an unlabeled target domain as follows. Formally, the labeled source domain is defined as:

$$\mathcal{D}_s = \{(x_i^S, y_i^S)\}_{i=1}^{N_s} \tag{1}$$

where x_i^S is a page image and y_i^S is its set of ground-truth bounding boxes and labels, and N_s is the number of labeled source pages.

The unlabeled target domain is expressed as:

$$\mathcal{D}_t^U = \{x_j^T\}_{j=1}^{N_t} \tag{2}$$

where x_j^T are page images without annotations, and N_t is the number of unlabeled target pages. When explicitly used, a small subset of manually annotated target data is denoted as

$$\mathcal{D}_t^L = \{(x_k^T, y_k^T)\}_{k=1}^{N_t^L}. \tag{3}$$

where N_t^L is the number of labeled target pages in this supervised subset. Here, the subscript $t \in \{A, B\}$ specifies the target newspaper: A corresponds to *Adresseavisen* and B corresponds to *iTromsø* (iTromsø, 2023). The following subsections detail the methodologies, grouped by supervision level (Sections 4.1, 4.2).

4.1. Unsupervised domain adaptation with pseudo-labeling

A key challenge in domain adaptation is the lack of labeled data from the target domain. Without annotations, directly training or fine-tuning a detector on the target pages is infeasible. To address this, in the first setting, we adopt an unsupervised strategy in which pseudo-labels with automatically generated annotations serving as a substitute for manual labeling. Our adaptation pipeline, illustrated in Fig. 3, proceeds as follows:

1. **Train the source model (AdVision).** First, the base YOLO detector f_θ, denoted as AdVision, is trained on the labeled

source dataset $\mathcal{D}_s$, minimizing the detection loss $\mathcal{L}_{\text{det}}$, obtaining a source trained AdVision model f_{θ_S}:

$$\theta_S = \arg\min_{\theta} \frac{1}{|\mathcal{D}_s|} \sum_{(x_i^S, y_i^S) \in \mathcal{D}_s} \mathcal{L}_{\text{det}}\big(f_\theta(x_i^S),\, y_i^S\big). \tag{4}$$

where $\mathcal{L}_{\text{det}}$ is the standard YOLOv8 detection loss, consisting of a weighted sum of bounding-box regression loss (CIoU), objectness loss, and classification loss.

2. **Generate pseudo-labels on the target domain.** Next, the trained AdVision model is applied to each unlabeled target image $x_j^T \in \mathcal{D}_t^U$ to produce candidate annotations:

$$\hat{y}_j^T = f_{\theta_S}(x_j^T) = \{(b_{j_m}, c_{j_m}, s_{j_m})\}_{j_m=1}^{j_M}, \tag{5}$$

where each predicted instance is represented as a tuple $(b_{j_m}, c_{j_m}, s_{j_m})$ consisting of a bounding box b_{j_m}, a class label c_{j_m}, and an associated confidence score $s_{j_m} \in [0, 1]$. Here, index $j_m (1 \le j_m \le j_M)$ denotes the *m-th* detection instance out of j_M detection instances in input image j. We apply two distinct confidence thresholds: a high cutoff τ_{adv} for keeping Ad pseudo-boxes, and a low cutoff τ_{null} for declaring a page as Null. For a target image x_j^T, the Ad pseudo-label set is

$$\tilde{y}_j^{T,adv} = \{(b_{j_m}, c_{j_m}) \mid j_m \in \{1, \ldots, j_M\}, s_{j_m} \ge \tau_{adv}\} \tag{6}$$

The Ad-pseudo subset P_{adv} consists of all target pages with at least one detection above the confidence threshold:

$$P_{adv} = \left\{(x_j^T, \tilde{y}_j^{T,adv}) \,\middle|\, j \in \{1, \ldots, N_T\},\ \tilde{y}_j^{T,adv} \ne \varnothing\right\}. \tag{7}$$

Conversely, A page is pseudo-labeled as Null if it contains no detections even at the permissive cutoff τ_{null}:

$$\tilde{y}_j^{T,null} = \{(b_{j_m}, c_{j_m}) \mid j_m \in \{1, \ldots, j_M\},\ s_{j_m} \ge \tau_{null}\}. \tag{8}$$

$$P_{null} = \left\{(x_j^T, \tilde{y}_j^{T,null}) \,\middle|\, j \in \{1, \ldots, N_t\},\ \tilde{y}_j^{T,null} = \varnothing\right\}. \tag{9}$$

For clarity, we denote these pseudo-labeled subsets by $\mathcal{D}_t^{\text{PL}\alpha}$. Here, $t \in \{A, B\}$ refers to the target newspapers, with A corresponding to Adresseavisen and B to iTromsø. The parameter α specifies the thresholding strategy (e.g., $\alpha = 96$ for $\tau_{\text{adv}} = 0.96$, $\alpha = 5$ for $\tau_{\text{null}} = 0.05$). Exact threshold values are detailed in the Results section. It is important to distinguish between pseudo-labels subsets and ground-truth dataset statistics. Specifically, P_{adv} and P_{Null} represent pseudo-labeled target pages identified by the source-trained model. In contrast, N_{adv} and N_{Null} in Table 1 in Section 5.1 denote the numbers of ground-truth Ad and Null page available in labeled datasets. Applying our pseudo-labeling procedure typically results in $|P_{\text{adv}}|$ and $|P_{\text{Null}}|$ that differ from the ground-truth counts, depending on the selected confidence thresholds.

3. **Form the combined training set (as illustrated by the orange arrow in Fig. 3).** $\mathcal{D}_{\text{combined}}$ is the combination of the source data and the chosen pseudo-labeled target data:

$$\mathcal{D}_{\text{combined}} = \mathcal{D}_s \cup \begin{cases} P_{adv} \\ P_{Null} \\ P_{adv} \cup P_{Null} \end{cases} \tag{10}$$

Depending on the strategy, $\mathcal{D}_{\text{combined}}$ augments the source set with only pseudo-labeled Ad pages (P_{adv}), only pseudo-labels Null pages (P_{Null}), or both.

4. **Train the adapted model (AdAPT).** A new YOLO detector, denoted as AdAPT f_θ, is trained on $\mathcal{D}_{\text{combined}}$:

$$\theta_{AdAPT} = \arg\min_{\theta} \frac{1}{|\mathcal{D}_{\text{combined}}|} \sum_{(x,y) \in \mathcal{D}_{\text{combined}}} \mathcal{L}_{\text{det}}\big(f_\theta(x),\, y\big). \tag{11}$$

This adapted model produces the final predictions on the target domain.

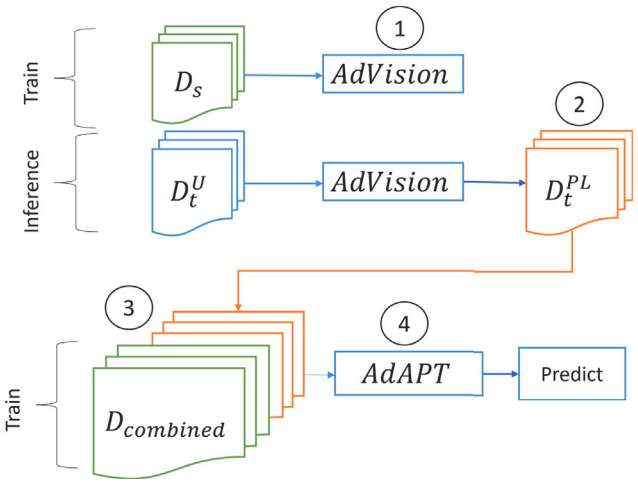


Fig. 3. Illustration of the unsupervised domain adaptation with pseudo-labeling. The source model AdVision (Sayyad et al., 2025) (YOLOv8-based (Reis et al., 2023)) is first trained on labeled source data $\mathcal{D}_s$ (Stage 1). It is then applied to unlabeled target pages $\mathcal{D}_t^U$ to generate pseudo-labels data $\mathcal{D}_t^{PL}$ (Stage 2). The pseudo-labeled target pages are combined with the source set to form $\mathcal{D}_{\text{combined}}$ (Stage 3), which is used to train the adapted detector AdAPT(Stage 4). The adapted model is then evaluated on the target domain (Predict).

4.2. Limited supervised adaptation

In the second setting, unlike the previous pseudo-label based adaptation, this approach does not rely on any automatically generated target-domain labels. Instead, we investigate the effect of adding a limited amount of manually annotated target samples to the training set. The objective is to quantify how much labeled target data is needed to effectively adapt the model when introducing a new newspaper.

We define a small subset of labeled target data as:

$$\mathcal{D}_t^L = \{(x_k^T, y_k^T)\}_{k=1}^{N_t^L}, \tag{12}$$

where x_k^T is a target-domain image and y_k^T its corresponding manual annotation. To assess the trade-off between annotation effort and performance gain, we experiment with adding $N_t^L \in \{67, 134, 334, 558\}$ labeled target samples, corresponding to approximately 1.0%, 2.0%, 5.0%, and 8.3% of the source training set ($N_s = 6696$). The smallest setting (67 samples) was chosen to simulate a low-effort adaptation scenario requiring minimal manual annotation.

For each N_t^L level, a new YOLO-based model is trained from scratch using the combined dataset

$$\mathcal{D}_{\text{train}} = \mathcal{D}_s \cup \mathcal{D}_t^L, \tag{13}$$

where $\mathcal{D}_s$ is the original labeled source dataset. The resulting model is evaluated on the target-domain test set, and accuracy gains are reported for each N_t^L. This analysis provides insight into the minimum proportion of labeled target data required to achieve satisfactory adaptation when adding a new newspaper domain.

5. Experiments

In this section, we present the experimental setup including datasets, training configuration, and evaluation metrics used to evaluate our proposed adaptation strategies.

Table 1
Dataset statistics: number of Null pages , Ad pages , and total pages used in the experiments. Subscripts L and U denote labeled and unlabeled splits, respectively.

Dataset	N_{Null}	N_{adv}	N_{Total}	Dataset period	Newspaper name	Type
D_s	3348	3348	6696	04/2022–03/2023	Berlingske, Sydsvenskan, UK Metro	Source train
D_A^L	2676	2573	5249	04/2022–03/2023	Adresseavisen	Target train
	283	275	558	04/2022–03/2023		Target test
D_A^U	–	–	12 202	01/2024–09/2024	Adresseavisen	Target inference (unlabeled)
D_B^L	139	135	274	05/2023	iTromsø	Target test
D_B^U	–	–	37 292	01/2024–09/2024	iTromsø	Target inference (unlabeled)

Table 2
Pseudo-labeled (PL) datasets obtained from the two target domains D_A^U (Adresseavisen) and D_B^U (iTromsø). Ad pages use confidence thresholds $\tau_{\text{adv}} \in \{0.96, 0.97\}$, producing D_*^{PL96} and D_*^{PL97}, respectively. Null pages use $\tau_{\text{null}} = 0.05$, producing D_*^{PL5}.

Dataset	Threshold	Pseudo-Ad Pages	Pseudo-Null Pages	Newspaper name	Type
D_B^{PL96} (iTromsø)	$\tau_{\text{adv}} = 0.96$	1964	–	iTromsø	Pseudo-labeled
D_A^{PL96} (Adresseavisen)	$\tau_{\text{adv}} = 0.96$	2112	–	Adresseavisen	Pseudo-labeled
D_B^{PL97} (iTromsø)	$\tau_{\text{adv}} = 0.97$	571	–	iTromsø	Pseudo-labeled
D_A^{PL97} (Adresseavisen)	$\tau_{\text{adv}} = 0.97$	893	–	Adresseavisen	Pseudo-labeled
D_B^{PL5} (iTromsø)	$\tau_{\text{null}} = 0.05$	–	2000	iTromsø	Pseudo-labeled
D_A^{PL5} (Adresseavisen)	$\tau_{\text{null}} = 0.05$	–	1279	Adresseavisen	Pseudo-labeled

5.1. Dataset

Our study uses digitized newspaper datasets spanning multiple domains. Each domain corresponds to a different publication with its own layout, typography, and editorial style. Our experiments focus on a set of annotated newspapers as source and target benchmarks, complemented by much larger collections of unannotated data that were used for pseudo-labeling. The annotated benchmarks include three source newspapers: Berlingske (Denmark) (Berlingske, 2022–2023), Sydsvenskan (Sweden) (Sydsvenskan, 2022–2023), and UK Metro (England) (Metro, 2022–2023), combined to form the labeled source domain D_s. To assess cross-publisher generalization, we use Adresseavisen (Norway) (Adresseavisen, 2022–2023) as the target domain, with a small annotated subset D_A^L and a larger unlabeled set D_A^U. In addition, we collected unannotated pages from iTromsø (Norway) (iTromsø, 2023) D_B^U to enable further pseudo-labeling experiments.

We relied on Adresseavisen with both labeled (D_A^L) and unlabeled splits (D_A^U) which enabled controlled evaluation of supervised adaptation with limited labeled target data alongside with unsupervised pseudo-labeling. In contrast, iTromsø (D_B^U) was used as a purely unlabeled domain during adaptation, with its small annotated subset (D_B^L) reserved for testing and analysis. Although both are Norwegian regional newspapers, they differ in layout density, typography, and advertisement frequency, providing complementary test conditions.

Table 1 summarizes the datasets used in this study. The final column ('Type') indicates the role of each dataset split in our experiments (source training, target training, target testing, or target inference for pseudo-labeling). We denote a Null page as a newspaper image without any Ads, and an Ad page as a page containing at least one Ad. In the table, N_{Null} and N_{adv} refer to the number of pages of each type, and N_{Total} is the total number of pages. Table 2 reports the number of target pages that were pseudo-labeled as Ad pages under the high-confidence thresholds (PL96/PL97) and the number that were identified as Null pages under the cutoff threshold (PL5). These pseudo-labeled subsets are used in the adaptation experiments in Section 6.

5.2. Evaluation metrics

In publisher workflows and archival analytics, false positives and false negatives affect downstream use differently. False positives (classifying editorial content as an Ad) increase the estimated share of advertising and reduce the accuracy of editorial statistics. False negatives (missing true Ads) lower the estimated volume of advertising content, which can affect analyses of advertising activity and space allocation. We therefore evaluate models with four standard object-detection metrics: (i) mAP@50 (mean Average Precision at IoU = 0.5); (ii) mAP@50–95, averaged over IoU = 0.50:0.95 in 0.05 steps; (iii) Precision = $\frac{TP}{TP+FP}$, where TP = true positives and FP = false positives; and (iv) Recall = $\frac{TP}{TP+FN}$, where FN = false negatives.

Precision penalizes false positives (controlling inventory inflation risk), while Recall penalizes false negatives (controlling revenue undercount risk). We select models by mAP@50, applying statistical tests on mAP@50 scores to confirm significance. We prioritize mAP@50 because Ad boxes follow layout/design boundaries where IoU = 0.5 is an accepted tolerance, while it aggregates performance across confidence thresholds (usable operating points). Precision and Recall are also reported so practitioners can choose deployment thresholds consistent with their FP/FN tolerance.

5.3. Experimental setup

All experiments were conducted on an NVIDIA GeForce RTX 3090 GPU (with 24 GB VRAM) with fixed random seed for reproducibility. We employed the YOLOv8 architecture (Jocher et al., 2023) trained from scratch for 150 epochs. Preliminary, we trained different YOLOv8 variants (n, s, m) on our dataset and observed only marginal performance differences; therefore, we selected the lightweight YOLOv8n to reduce computation and training time without sacrificing accuracy.

The final training configuration was determined through hyperparameter tuning in early experiments, where we tested different optimization and augmentation settings to balance accuracy, training stability, and efficiency. Table 3 summarizes the training configuration and Table 4 lists the data augmentations. The hyperparameters in Table 3 were determined via a structured search on the labeled source domain D_s. We defined initial ranges centered on the approximate defaults of the YOLOv8 (e.g., learning-rate 0.001, momentum ≈ 0.937, weight-decay 0.0005) and explored learning rate $\in [1\times10^{-3}, 5\times10^{-3}]$, momentum $\in$ [0.85, 0.95], weight-decay $\in [1\times10^{-4}, 5\times10^{-4}]$, using a cosine-decay schedule with warm-up. The optimum setting (learning rate = 2.74×10^{-3}, momentum = 0.8818, weight-decay = 3.6×10^{-4}) achieved the most stable convergence and highest validation mAP on D_s. The warm-up duration of 4.93 epochs was selected by the search and represents a continuous transition from warm-up to full scheduling. This fractional epoch value provided improved early-stage stability compared with integer warm-up steps.

The labeled source domain D_s combined Berlingske (DK), Sydsvenskan (SE), and UK Metro (UK). Unlabeled target domains were Adresseavisen and iTromsø, and their small labeled subsets, D_A^L and

Table 3
Training configuration for all experiments.

Component	Setting
Model architecture	YOLOv8n (lightweight, chosen over s/m)/ YOLOv12n
Epochs	150
Batch size	16
Input resolution	608 × 608
Optimizer	SGD
Initial LR	2.74×10^{-3}
LR schedule	Cosine decay (lrf = 0.00839)
Momentum	0.8818
Weight decay	3.6×10^{-4}
Warmup epochs	4.93 (momentum = 0.7061)
Loss weights	box = 6.06, cls = 0.35, DFL = 1.51

Table 4
Data augmentation configuration.

Augmentation	Setting
HSV jitter	hue = 0.0098, sat = 0.4393, val = 0.2671
Translation	0.086
Scaling	0.223
Horizontal flip	0.5
Mosaic	0.518
Disabled transforms	rotation, shear, perspective, mixup, copy–paste

D_B^L, used only for evaluation. For pseudo-labeling, Ad pages were derived from the high-confidence sets $D^{\text{PL96–97}}$, and Null pages were derived from D^{PL5}. Models were evaluated on the target-domain test sets using Precision, Recall, mAP@50, and mAP@50–95, reporting mean and standard deviation over five runs initialized with fixed random seeds.

To assess whether observed performance differences across training scenarios were statistically significant, we conducted a one-way analysis of variance (ANOVA) on the mAP@50 scores obtained from five independent training runs per configuration. ANOVA tests the null hypothesis that all group means are equal, and a significant result indicates that at least one scenario differs (Fisher, 1925). When ANOVA revealed significant variation, we applied Tukey's post-hoc Honestly Significant Difference (HSD) test (Tukey, 1949) to identify which pairs of scenarios differed at $\alpha = 0.05$. This procedure controls the family-wise error rate when performing multiple comparisons and is widely used in experimental evaluation.

6. Results

This section presents results on three aspects: the effect of adding limited labeled target data Fig. 6; the comparison of Ad and Null page pseudo-labeling strategies Tables 7, 6, and Fig. 5; and the influence of confidence thresholds on pseudo-label selection. Dataset statistics for these experiments are summarized in Table 1. In all experiments, models were trained on the combined source dataset (Berlingske, Sydsvenskan, UK Metro) and optionally augmented with pseudo-labeled Ad pages , pseudo-labeled Null pages , both, or a small subset of labeled target-domain pages. For our proposed approach, AdAPT, performance was evaluated on two target domains — D_A^L (Adresseavisen) and D_B^L (iTromsø). The limited supervised adaptation was conducted using D_A^L as the labeled target dataset, since only this domain provided a sufficiently large and balanced annotated subset for experiments. In the following experiments, we denote the threshold-specific pseudo-labeled datasets as D_t^{PL96}, D_t^{PL97}, and D_t^{PL5}, corresponding respectively to $\tau_{\text{adv}} = 0.96$, $\tau_{\text{adv}} = 0.97$, and $\tau_{\text{null}} = 0.05$. Before proceeding to the domain adaptation experiments, we compared three YOLOv8 variants (n, s, m) to assess the trade-off between model capacity, computational cost, and detection accuracy. All variants were trained with the same configuration specified in Tables 3 and 4. Each model was trained using five-fold cross-validation with a fixed random seed (1), and performance was measured on the labeled target-domain test sets D_A^L and D_B^L. Table 5 summarizes the resulting mean ± standard-deviation metrics, showing that YOLOv8n achieves accuracy comparable to YOLOv8s and YOLOv8m on both D_A^L and D_B^L. However, YOLOv8n trains roughly *2–3× faster* than YOLOv8m, making it more efficient for repeated adaptation runs. Since AdAPT requires multiple pseudo-labeling and fine-tuning cycles, this efficiency consideration is important. For this reason, YOLOv8n was selected as the backbone for all subsequent adaptation experiments.

6.1. AdAPT

In practice, we employ two distinct operating points for pseudo-labeling. For Ad pseudo-labels, we use a high threshold $\tau_{\text{adv}} = 0.97$, denoted D_t^{PL97}. We found that thresholds above 0.97 produced less than ten pseudo-labeled pages, which limited their utility, whereas D_t^{PL97} provided sufficient samples. In addition to D_t^{PL97}, we also evaluate D_t^{PL96}as a sensitivity check within the same high-confidence regime. D_t^{PL96}slightly relaxes the cutoff to admit a few more Ad pseudo-labels. For identifying Null pseudo-labeled pages, a much lower threshold is required: a page is considered Null if it contains no detections even under this permissive cutoff. To decide on the cutoff, we conducted a validation experiment on the small labeled target subset D_A^L, sweeping τ values up to 0.40 and measuring page-level precision/recall, as shown in Fig. 4.

To assess the reliability of Null page selection across different confidence thresholds, predictions were evaluated against the manually annotated target subset D_A^L. Using this labeled data, Precision, Recall, and F1-score were computed for Null page identification under varying thresholds. As illustrated in Fig. 4, increasing the confidence threshold leads to higher recall but a decline in precision. At higher thresholds, fewer detections exceed the confidence cutoff, causing more pages to be classified as Null pages . Consequently, a larger portion of true Null pages is correctly identified (increasing recall), but some pages containing low-confidence Ads are mistakenly included as null, which reduces precision. For Null pseudo-labeling, maximizing precision is more critical than recall, since even a small proportion of mislabeled Ad pages can introduce considerable noise into the training set. Therefore, D_t^{PL5} ($\tau_{\text{null}} = 0.05$) was selected as the optimal confidence threshold, providing near-perfect precision and ensuring that pseudo-labeled Null pages serve as a reliable source of negative supervision.

After establishing $\tau_{\text{null}} = 0.05$ as the optimal confidence threshold, we next investigated how pseudo-labeled data affected model adaptation across target newspapers. Mean performance and standard deviations, computed over five independent runs, are reported in Tables 6, and 7. We evaluated three adaptation strategies: (i) Ad-only pseudo-labels, filtered with a high confidence threshold ($D_t^{\text{PL96–97}}$); (ii) Null-only pseudo-labels, selected with a low threshold (D_t^{PL5}); and (iii) the combination of Ad and Null pseudo-labels. Across both domains, adding D_t^{PL5} to the source data achieved the highest gains. Adding D_A^{PL5} increased mAP@50 from 0.69 to 0.79, while adding D_B^{PL5} raised performance from 0.58 to 0.66, as reported in Tables 6 and 7. In contrast, using D_A^{PL96} reduced accuracy to 0.60 mAP@50 and D_B^{PL96} to 0.55, both performing below the source-only baseline.

While these mean differences suggest a strong benefit from Null-based pseudo-labeling, statistical verification was necessary to confirm that the observed gains were not due to random variation across runs. Therefore, a one-way analysis of variance (ANOVA) was conducted to compare the six training configurations for each domain. For the D_A^L, the ANOVA revealed significant differences among the strategies ($F(5, 24) = 22.71$, $p < 10^{-7}$). Subsequent Tukey HSD post-hoc comparisons confirmed that adding D_A^{PL5} significantly outperformed both the baseline and the Ad-only pseudo-labeling configurations (adding D_A^{PL96}, D_A^{PL97}) , with mean mAP@50 improvements of approximately 14%–19% over the baseline ($p < 0.01$). In contrast, adding D_A^{PL96} and D_A^{PL97} strategies did not differ significantly from the baseline ($p > 0.05$).

Table 5
Performance of YOLOv8n, YOLOv8s, and YOLOv8m on the $\mathcal{D}_B^L$ (iTromsø) and $\mathcal{D}_A^L$ (Adresseavisen) test sets. Results are reported as mean ± std across five folds (seed = 1).

Model	Test dataset	Training Time (h)	P	R	mAP@50	mAP@50–95
YOLOv8n	$\mathcal{D}_B^L$	**0.59 ± 0.005**	**0.56 ± 0.053**	**0.63 ± 0.083**	**0.61 ± 0.036**	0.53 ± 0.029
YOLOv8n	$\mathcal{D}_A^L$	**0.59 ± 0.005**	**0.62 ± 0.061**	0.64 ± 0.037	**0.69 ± 0.036**	0.60 ± 0.025
YOLOv8s	$\mathcal{D}_B^L$	0.88 ± 0.004	0.53 ± 0.082	0.62 ± 0.054	0.58 ± 0.043	0.51 ± 0.021
YOLOv8s	$\mathcal{D}_A^L$	0.88 ± 0.004	0.59 ± 0.060	0.63 ± 0.037	0.65 ± 0.055	0.57 ± 0.041
YOLOv8m	$\mathcal{D}_B^L$	1.72 ± 0.004	0.55 ± 0.085	0.59 ± 0.018	0.58 ± 0.036	0.51 ± 0.027
YOLOv8m	$\mathcal{D}_A^L$	1.72 ± 0.004	0.58 ± 0.083	**0.67 ± 0.044**	0.68 ± 0.054	0.58 ± 0.036

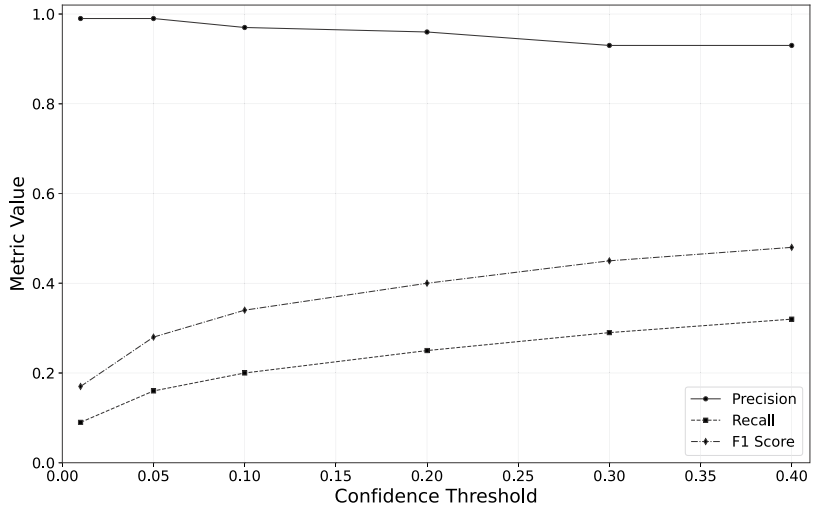
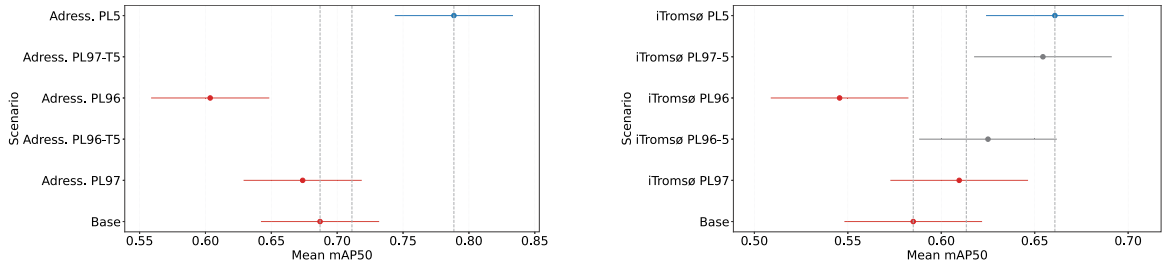


Fig. 4. Precision, Recall, and F1-score for Null page identification under varying confidence thresholds.



(a) Results for $\mathcal{D}_A^L$.

(b) Results for $\mathcal{D}_B^L$.

Fig. 5. Tukey's Honestly Significant Difference (HSD) post-hoc analysis of mean mAP@50 across adaptation scenarios using AdVision (source-only baseline) and AdAPT (pseudo-labeled variants). Each point indicates the mean mAP@50 over five independent runs, and horizontal bars represent 95% confidence intervals based on Tukey's pooled variance. Dashed vertical lines denote the baseline mean, overall mean, and the best-performing configuration. Results are presented for (a) $\mathcal{D}_A^L$ (Adresseavisen) and (b) $\mathcal{D}_B^L$ (iTromsø).

For the $\mathcal{D}_B^L$, overall performance differences were also statistically significant ($F(5, 24) = 13.36$, $p < 10^{-5}$). The subsequent Tukey HSD pairwise analysis identified $\mathcal{D}_B^{\text{PL5}}$ as the best-performing configuration, whereas $\mathcal{D}_B^{\text{PL96}}$ and $\mathcal{D}_B^{\text{PL97}}$ produced no statistically reliable gains relative to the baseline. These results confirm that the improvement observed with the Null-based pseudo-labeling strategy is consistent and unlikely to result from random variation across training runs. Full Tukey results are summarized in Fig. 5, where only $\mathcal{D}_B^{\text{PL5}}$'s confidence interval lies entirely above the baseline, indicating its significant and stable performance advantage.

6.2. Impact of limited labeled target data

In practical applications, obtaining labeled target-domain data is costly and time-consuming, making it important to understand how much annotation is needed to achieve reliable adaptation. To determine how much labeled target data is required for reliable Ad detection performance in the target newspaper, we analyzed mAP@50 scores as progressively larger subsets of $\mathcal{DA}^L$ were added to the source dataset. As illustrated in Fig. 6, the baseline model trained only on source data achieved 0.69 mAP@50 on $\mathcal{D}_A^L$. Adding 134 labeled pages to the source dataset increased the score to 0.84, and extending it to 335 labeled pages raised it to 0.85. Performance continued to improve with additional labeled data, reaching 0.86 mAP@50 at 558 pages. For comparison, a model trained solely on 134 pages from $\mathcal{D}_A^L$ (without source data) reached only 0.67 mAP@50, close to the baseline.

7. Discussion

Our experiments provide several insights into adapting Ad detectors across diverse newspaper domains. Below we summarize we discuss the key findings.

Table 6
AdAPT results evaluated on the $\mathcal{D}_B^L$ (iTromsø) test set. Each model was trained on the same labeled source dataset $\mathcal{D}_s$, optionally augmented with pseudo-labeled (PL) target data. Columns "Ad PL" and "Null PL" indicate which pseudo-labeled subsets were included: : $\mathcal{D}_B^{PL96}$, $\mathcal{D}_B^{PL97}$ (Ad pages) or $\mathcal{D}_B^{PL5}$ (Null pages). A value of "—" means that no additional PL data of that type was included in the training set.

Ad PL	Null PL	P	R	mAP@50	mAP@50_95
–	–	0.58 ± 0.032	0.56 ± 0.055	0.58 ± 0.029	0.51 ± 0.029
$\mathcal{D}_B^{PL96}$	–	0.55 ± 0.021	0.56 ± 0.043	0.55 ± 0.029	0.48 ± 0.026
$\mathcal{D}_B^{PL97}$	–	0.57 ± 0.034	0.61 ± 0.042	0.61 ± 0.010	0.53 ± 0.008
–	$\mathcal{D}_B^{PL5}$	0.61 ± 0.058	**0.66 ± 0.072**	**0.66 ± 0.038**	**0.58 ± 0.033**
$\mathcal{D}_B^{PL97}$	$\mathcal{D}_B^{PL5}$	0.60 ± 0.048	0.64 ± 0.045	0.65 ± 0.007	0.57 ± 0.006
$\mathcal{D}_B^{PL96}$	$\mathcal{D}_B^{PL5}$	**0.64 ± 0.028**	0.62 ± 0.047	0.62 ± 0.011	0.55 ± 0.011

Table 7
AdAPT results evaluated on the $\mathcal{D}_A^L$ (Adresseavisen) test set. All models were trained on a base combined set of source data $\mathcal{D}_s$, optionally augmented with pseudo-labels (PL) data. Columns "Ad PL" and "Null PL" indicate which pseudo-labeled subsets were included: $\mathcal{D}_A^{PL96}$, $\mathcal{D}_A^{PL97}$ (Ad pages) or $\mathcal{D}_A^{PL5}$ (Null pages). A value of "—" means that no additional PL data of that type was included in the training set.

Ad PL	Null PL	P	R	mAP@50	mAP@50_95
–	–	0.60 ± 0.039	0.68 ± 0.037	0.69 ± 0.044	0.59 ± 0.035
$\mathcal{D}_A^{PL96}$	–	0.52 ± 0.018	**0.72 ± 0.034**	0.60 ± 0.025	0.52 ± 0.022
$\mathcal{D}_A^{PL97}$	–	0.56 ± 0.081	0.64 ± 0.052	0.67 ± 0.033	0.58 ± 0.031
–	$\mathcal{D}_A^{PL5}$	**0.75 ± 0.038**	0.72 ± 0.044	**0.79 ± 0.030**	**0.69 ± 0.023**
$\mathcal{D}_A^{PL97}$	$\mathcal{D}_A^{PL5}$	0.71 ± 0.018	0.69 ± 0.006	0.76 ± 0.007	0.67 ± 0.007
$\mathcal{D}_A^{PL96}$	$\mathcal{D}_A^{PL5}$	0.67 ± 0.025	0.69 ± 0.030	0.75 ± 0.021	0.65 ± 0.020

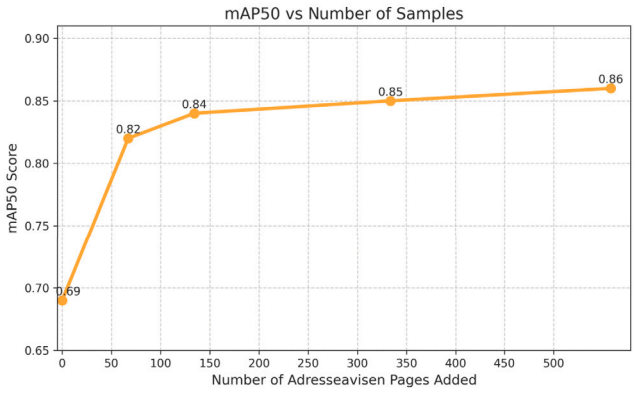


Fig. 6. Effect of labeled target data on model adaptation. The figure shows mAP@50 scores on the target domain $\mathcal{D}_A^L$ (Adresseavisen) as progressively larger subsets of labeled pages are added to the source dataset $\mathcal{D}s$.

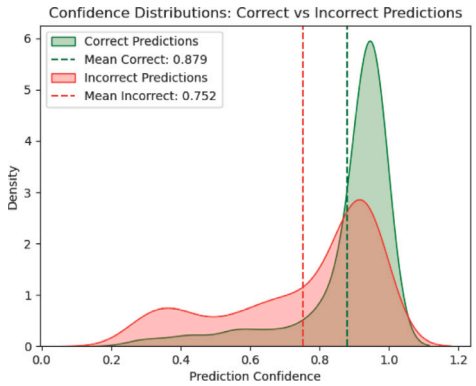
Pseudo-labeling strategies: The effect of pseudo-labeling differs between Ad pages and Null pages. Ad-only pseudo-labels ($\mathcal{D}_t^{PL96}$, $\mathcal{D}_t^{PL97}$) selected with strict confidence thresholds did not improve accuracy and sometimes reduced it, as shown in Tables 7, 6. The main issue is that correct and incorrect detections receive overlapping confidence scores. Figs. 7(a) for $\mathcal{D}_A^L$ and 7(b) for $\mathcal{D}_B^L$ show confidence distributions. Correct boxes have on average higher confidence (about 0.88), but incorrect ones also reach high values (0.73–0.75). Because kernel density estimation (KDE) was used for visualization, the curves extend slightly beyond the theoretical [0,1] interval; this is a smoothing artifact and does not imply prediction confidences greater than 1. The overlap between correct and incorrect detections means that a strict cutoff removes many true Ads together with false positives, leaving supervision that is both noisy and biased.

In contrast, add $\mathcal{D}_t^{PL5}$ consistently improved mAP and precision across both target domains, as shown in Tables 7, 6. Null pages are abundant in newspaper archives, and when used as pseudo-labels they provide reliable negative supervision. This helps the detector learn when no Ad is present, reduces false positives, and improves calibration against background variation. Combining Ad and Null pseudo-labels did not lead to additional improvements. The noise in Ad pseudo-labels outweighed any incremental signal once the model was already well regularized by Null pages.

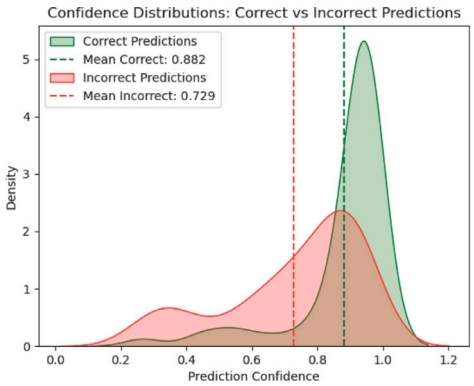
These findings suggest that, under domain shift, pseudo-labels representing negative samples (i.e., pages confidently identified as containing no Ads) are more effective than noisy positive detections. Null pages cover a broader portion of the feature space and therefore help constrain the decision boundary, improving generalization. Ad-based pseudo-labeling could become more effective if coupled with confidence calibration or iterative re-labeling schemes that better separate true and false detections.

Labeled target data boosts adaptation: Pseudo-labeling provides a more scalable but less precise alternative compared to limited target adaption. Adding $\mathcal{D}_t^{PL5}$ achieves performance close to that obtained with a small labeled subset of the target domain, without requiring manual annotation. However, a small annotation budget provide large adaptation gains when combined with diverse source data, as shown in Fig. 6. In practice, handling the entire labeled target domain for adaptation would be costly and often unnecessary, since the goal is to enable low-touch updates rather than full retraining on large target sets. Therefore, we selected a small, representative labeled subset of $\mathcal{D}_t^L$ to balance adaptation accuracy with annotation and compute efficiency. Labeled target pages supply precise supervision that help the model adjust its decision boundaries to the target layout and correct residual biases from source domains. In contrast, training solely on a limited target set performs comparably to the baseline, indicating that labeled data are most effective when used to complement diverse source supervision rather than replace it.

Validation on iTromsø variants We further evaluate on three iTromsø test variants: the original full test set ($\mathcal{D}_B^L$), a version capped at two Ads per page ($\mathcal{D}_B^{L\text{-Max2}}$), and a version capped at one Ad per page ($\mathcal{D}_B^{L\text{-Max1}}$); see Fig. 8. Fig. 8(a) shows the variation in mAP@50 across the evaluated datasets. Adding $\mathcal{D}_B^{PL5}$ (Null-only) gives the strongest overall results, rising from about 0.66 on the $\mathcal{D}_B^L$ to 0.73 on $\mathcal{D}_B^{L\text{-Max2}}$ and 0.80 on ($\mathcal{D}_B^{L\text{-Max1}}$). In contrast, $\mathcal{D}_B^{PL97}$ (Ad-only) improves more modestly, reaching only 0.61, 0.68, and 0.74 across the three splits. Adding both Ad and Null pages $\mathcal{D}_B^{PL97\text{–}5}$, lies between them, at roughly



(a) First target, $\mathcal{D}_A^L$ (Adresseavisen). Correct detections are typically more confident (mean ≈ 0.882), but incorrect detections still receive moderately high confidence (mean ≈ 0.729), yielding noisy Ad pseudo-labels at strict thresholds.



(b) Second target, $\mathcal{D}_B^L$ (iTromsø). A similar overlap occurs (means ≈ 0.879 vs. ≈ 0.752), confirming no single threshold cleanly separates correct from incorrect Ad detections.

Fig. 7. Confidence distributions of correct vs. incorrect predictions (density). Overlap at high confidence explains why Ad-only pseudo-labels are weak while Null-only augmentation is robust.

0.65, 0.72, and 0.79. This shows that while adding high-confidence Ad pseudo-labels gives some benefit, they are not as consistently reliable as Null page. The largest gap appears on the $\mathcal{D}_B^L$, where the more diverse layouts expose the noise in Ad pseudo-labels most clearly.

Across all training configurations, performance improves monotonically from $\mathcal{D}_B^L$ to $\mathcal{D}_B^{L\text{-Max2}}$ and then to $\mathcal{D}_B^{L\text{-Max1}}$, indicating that the models perform best when the test distribution matches the simpler one-ad-per-page scenario. Looking at recall and precision separately in Figs. 8(b), 8(c) clarifies mAP50 behavior. Adding $\mathcal{D}_B^{\text{PL5}}$ to source data consistently raises precision, climbing from around 0.61 to 0.71, which reflects a clear reduction of false positives on target-domain backgrounds. Adding $\mathcal{D}_B^{\text{PL97-5}}$ achieves the highest recall on the $\mathcal{D}_B^{L\text{-Max1}}$ split (around 0.81), as most Ad pseudo-labels collected at strict thresholds come from pages with exactly one Ad; when the test set matches this distribution, the model can recover nearly all Ads. However, this recall gain comes at the cost of reduced precision, which explains why the overall mAP does not surpass adding $\mathcal{D}_B^{\text{PL5}}$. By contrast, $\mathcal{D}_B^{\text{PL97}}$ alone suffers both lower recall and weaker precision, showing that sparse and noisy Ad pseudo-labels degrade rather than strengthen the adapted model.

These patterns highlight two main issues. First, the Ad pseudo-label pool is highly biased: strict thresholds preserve mostly single-Ad pages , while multi-Ad pages are often discarded. This creates a distribution mismatch that shows up as weaker performance on the $\mathcal{D}_B^L$. Second, because the confidence distributions of correct and incorrect boxes overlap substantially (see Fig. 7), lowering the threshold to collect more Ads introduces many false positives, thereby reducing precision and mAP50. In contrast, Null pages are abundant and reliable, and incorporating them calibrates the detector against spurious background patterns in the target newspapers, directly improving precision.

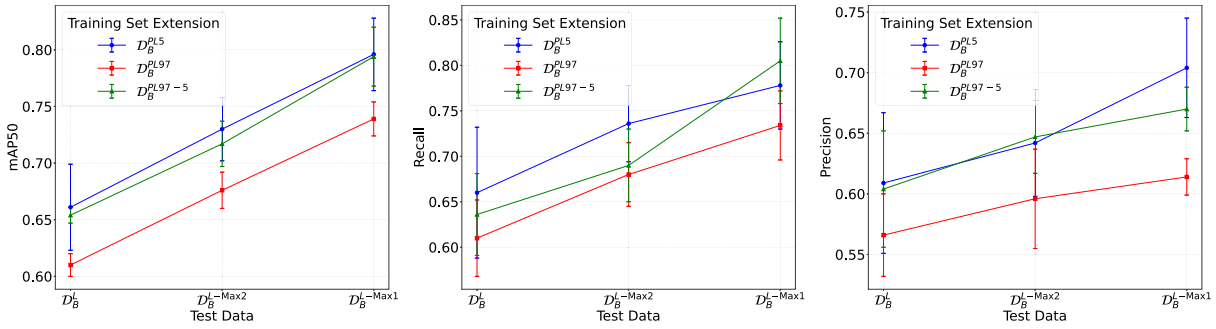
These experiments confirm that Null page pseudo-labels are the most stable source of adaptation gains, while Ad-page pseudo-labels remain a bottleneck. To move beyond this limitation, future work should focus on improving the quality of Ad pseudo-labels—for example by better pseudo-label filtering. Once Ad-page supervision becomes as reliable as Null page, combining the two should surpass Null-only augmentation, especially on denser target layouts. Until then, Null-only pseudo-labeling remains the most effective strategy.

Base detector choice (YOLOv8 vs. YOLOv12): To verify that the choice of base detector does not bias results, we additionally compared YOLOv8 and the newer YOLOv12 trained on the same labeled source data and evaluated on the source and target newspapers. Table 8 summarizes the results averaged over five folds. For this comparison, we fixed the random seed to 5 so that only variability due to data folds was assessed. Performance remained highly similar across both versions: YOLOv8 achieved higher mAP@50 on Adresseavisen (0.712 vs. 0.694) and comparable performance on iTromsø (0.572 vs. 0.618 for YOLOv12). These differences are small relative to the observed variation across folds and do not suggest a consistent performance advantage for YOLOv12. We observed nearly identical performance between YOLOv8 and YOLOv12, further reinforcing that detector choice does not materially influence the baseline. We therefore retain YOLOv8, consistent with the AdVision baseline (Sayyad et al., 2025), as a stable and representative architecture for advertisement detection. Moreover, even the newer YOLOv12, when trained without domain adaptation, does not outperform our Adapt method on the target domain, indicating that the main gains arise from adaptation rather than detector architecture.

AdAPT is model independent. It operates entirely at the pseudo-labeling and domain-adaptation stages and does not rely on any architectural features of the underlying detector. Consequently, AdAPT can be integrated with various detection architectures without modification. Future work may explore coupling AdAPT with newer detectors to further enhance cross-domain performance. Future work could explore integrating YOLOv12 within the Adapt framework to further enhance performance.

Implications and limitations: Our results establish Null-based pseudo-labeling as a low-cost, scalable strategy for cross-domain Ad detection. This is particularly advantageous in media archives where the majority of pages contain no Ads. in addition to cost reduction, such models can support downstream applications including large-scale media analysis, revenue modeling, and cultural heritage digitization. However, pseudo-labeled Ad pages remain a bottleneck. Confidence thresholds cannot cleanly separate correct from incorrect detections, resulting in pseudo-label sets that are not fully representative of real Ad-page variability. Future work should explore uncertainty-aware filters to make Ad pseudo-labels as robust as Null ones.

In this work, the confidence thresholds for pseudo-label selection were chosen empirically to ensure stable and high-precision supervision across domains. A promising future direction is to incorporate dynamic confidence calibration strategies that adjust thresholds according to model uncertainty. Recent studies, such as the Double Confidence Calibration Focused Distillation (DCCFD) framework (Fu et al., 2024), demonstrate how adaptive label smoothing and parameter-averaging mechanisms can jointly mitigate overconfidence and enhance stability



(a) mAP@50. (b) Recall. (c) Precision.

Fig. 8. Using our proposed approach, AdAPT, the source model was trained with three different target pseudo-labeled (PL) datasets to compare adaptation scenarios. Specifically, AdAPT was trained on the source data augmented with one of $\mathcal{D}_B^{\text{PL5}}$, $\mathcal{D}_B^{\text{PL97}}$, or $\mathcal{D}_B^{\text{PL97–5}}$. The adapted models were then evaluated on $\mathcal{D}_B^L$, $\mathcal{D}_B^{L\text{-Max2}}$ (up to two Ads per page), and $\mathcal{D}_B^{L\text{-Max1}}$ (single Ad per page). Panel (a) shows overall mAP@50; panels (b–c) present Recall and Precision. Error bars denote standard deviation across runs.

Table 8
Comparison between YOLOv8 and YOLOv12 trained on the labeled source domain and evaluated on target newspapers (mean ± standard deviation over five folds). Note: For this comparison, we fixed the random seed to 5 for both models to isolate variability due to data folds. Other experiments vary both seed and data fold and therefore use seeds 1–5.

Model	Test Dataset	Precision	Recall	mAP@50/mAP@50–95
YOLOv8	$\mathcal{D}_A^L$	**0.63** ± 0.080	0.67 ± 0.036	**0.71** ± 0.030/**0.61** ± 0.028
YOLOv12	$\mathcal{D}_A^L$	0.54 ± 0.050	**0.70** ± 0.073	0.69 ± 0.017/0.60 ± 0.017
YOLOv8	$\mathcal{D}_B^L$	0.51 ± 0.047	0.62 ± 0.061	0.57 ± 0.017/0.50 ± 0.017
YOLOv12	$\mathcal{D}_B^L$	**0.56** ± 0.031	**0.66** ± 0.047	**0.62** ± 0.024/**0.55** ± 0.023
YOLOv8n	$\mathcal{D}_s$	0.96 ± 0.018	0.97 ± 0.015	0.98 ± 0.006/0.97 ± 0.011
YOLOv12n	$\mathcal{D}_s$	0.96 ± 0.018	0.97 ± 0.012	0.98 ± 0.006/0.97 ± 0.010

during continual learning. Adopting similar uncertainty-aware calibration principles could further improve the reliability of pseudo-labeled Ad regions and complement the robust Null-based supervision presented in this study.

Although the present work focuses on Norwegian newspapers, the principles underlying AdAPT are not domain-specific and may be transferable to other document types that exhibit comparable layout variability, such as magazines, historical reports, or scientific journals. These materials often combine structured text regions with heterogeneous graphical or non-textual components (e.g., figures, tables, or advertisements), which similarly challenge model generalization under domain shift. Since AdAPT's confidence-guided pseudo-labeling relies primarily on reliable negative (Null) supervision rather than content-specific priors, its framework could be extended to such domains with minimal modification. Future research may investigate this broader applicability through multi-class document layout detection tasks.

8 Conclusion

This paper proposes AdAPT, a practical approach for adapting advertisement detection models to new newspaper domains with minimal annotation effort in the target. Our experiments showed that while small amounts of labeled target data can be highly effective, pseudo-labeling offers a practical unsupervised alternative. In particular, pseudo-labels Null pages representing Ad-free content proved to be the most reliable source of adaptation, improving generalization. By contrast, pseudo-labeled Ad pages alone were too noisy to drive consistent gains, and combining them with Null pages offered only marginal benefits. These findings highlight the potential of Confidence-guided pseudo-labeling, especially of Null examples, as a simple and scalable strategy for domain adaptation supporting large-scale media analysis in digital archives. Future work will extend this approach by investigating the impact of filtering noisy elements from Ad pages themselves, to further enhance cross-domain adaptation.

CRediT authorship contribution statement

Faeze Zakaryapour Sayyad: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – original draft, Visualization, Formal analysis. **Tobias Pettersson:** Industrial research, Data & algorithm development, Writing – review & editing. **Seyed Jalaleddin Mousavirad:** Writing – review & editing, Supervision. **Irida Shallari:** Writing – review & editing, Supervision. **Mattias O'Nils:** Writing – review & editing, Supervision.

Declaration of generative AI

During the preparation of this work the author(s) used openai in order to improve the language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Faeze Zakaryapour Sayyad reports financial support was provided by Mid Sweden University. Faeze Zakaryapour Sayyad reports a relationship with Knowledge Foundation (kks.se) within the Industrial graduate school Smart Industry Sweden, Media Research company that includes: employment and funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by The Knowledge Foundation (kks.se) within the Industrial graduate school Smart Industry Sweden and Media Research AB. The authors would also like to thank Oscar Berg for his valuable comments on this work.

Data availability

The authors do not have permission to share data.

References

Adresseavisen (2022–2023). Adresseavisen. Data extracted over the years 2022 and 2023. Accessed: 2023-04-26.

Alijani, S., Fayyad, J., & Najjaran, H. (2024). Vision transformers in domain adaptation and domain generalization: a study of robustness. *Neural Computing and Applications*, *36*(29), 17979–18007.

Almgren, K., Krishnan, M., Aljanobi, F., & Lee, J. (2018). AD or Non-AD: a deep learning approach to detect advertisements from magazines. *Entropy*, *20*(12), 982.

Almutairi, A. (2025). Newspaper elements detection and newspaper pages categorization using CNNs and transformers: Newspaper elements detection and newspaper pages categorization using CNNs and transformers. *International Journal on Document Analysis and Recognition (IJDAR)*, *28*(2), 225–237.

Almutairi, A., & Almashan, M. (2019). Instance segmentation of newspaper elements using mask R-CNN. In *2019 18th IEEE international conference on machine learning and applications* (pp. 1371–1375). IEEE.

Álvarez, J., Armenteros, J. C., Torrón, C., Ortega-Martín, M., Ardoiz, A., García, Ó., Arranz, I., Galdeano, Í., Garrido, I., Alonso, A., et al. (2024). RADIA–radio advertisement detection with intelligent analytics. arXiv preprint arXiv:2403.03538.

Bansal, A., Mukherjee, P., Joshi, D., Tripathi, D., & Singh, A. P. (2021). Multi-task learning for newspaper image segmentation and baseline detection using attention-based U-net architecture. In *Document analysis and recognition–ICDAR 2021 workshops: Lausanne, Switzerland, September 5–10, 2021, proceedings, part II 16* (pp. 440–454). Springer.

Berlingske (2022–2023). Berlingske. Data extracted over the years 2022 and 2023. Accessed: 2023-04-26.

Cai, M., Kezierbieke, J., Zhong, X., & Chen, H. (2024). Uncertainty-aware and class-balanced domain adaptation for object detection in driving scenes. *IEEE Transactions on Intelligent Transportation Systems*, *25*(11), 15977–15990.

Carvalho, P., Pereira, A., & Viana, P. (2021). Automatic tv logo identification for advertisement detection without prior data. *Applied Sciences*, *11*(16), 7494.

Chen, B., Chen, W., Yang, S., Xuan, Y., Song, J., Xie, D., Pu, S., Song, M., & Zhuang, Y. (2022). Label matching semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14381–14390).

Chen, Z., Luo, Y., Wang, Z., Baktashmotlagh, M., & Huang, Z. (2023). Revisiting domain-adaptive 3D object detection by reliable, diverse and class-balanced pseudo-labeling. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3714–3726).

Dhang, S., Zhang, M., & Dev, S. (2024). AdSegNet: a deep network to localize billboard in outdoor scenes. *Signal, Image and Video Processing*, *18*(10), 7221–7235.

Dhang, S., Zhang, M., & Dev, S. (2025). LaBINet-an approach for seamlessly integrating new advertisement into an existing scene. *IEEE Transactions on Artificial Intelligence*.

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.

Fu, Z., Wang, Z., Yu, C., Xu, X., & Li, D. (2024). Double confidence calibration focused distillation for task-incremental learning. *IEEE Transactions on Neural Networks and Learning Systems*, *36*(5), 9070–9083.

Guo, L., Zhang, Y., Liu, J., Liu, H., & Li, Y. (2025). Scale-consistent and temporally ensembled unsupervised domain adaptation for object detection. *Sensors*, *25*(1), 230.

He, Q., Xiao, S., Ye, M., Zhu, X., Neri, F., & Hou, D. (2023). Independent feature decomposition and instance alignment for unsupervised domain adaptation. In *Proceedings of the thirty-second international joint conference on artificial intelligence* (pp. 819–827).

He, R., Yang, J., & Qi, X. (2021). Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6930–6940).

iTromsø (2023). iTromsø. Data extracted for the dates 25-31 May 2023. Accessed: 2023-06-26.

Jain, P., Taneja, K., & Taneja, H. (2024). Advertisement image classification using deep learning with BERT: A novel approach exploiting textual. *Proceedings of Data Analytics and Management: ICDAM 2023, Volume 2*, *786*, 443.

Jain, P., et al. (2021). Convolutional neural network based advertisement classification models for online English newspapers. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(2), 1687–1698.

Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8.

Jonsson, P. (2022). A deep learning approach to advertisement detection in newspapers.

Kage, P., Rothenberger, J. C., Andreadis, P., & Diochnos, D. I. (2024). A review of pseudo-labeling for computer vision. ArXiv E-Prints, arXiv–2408.

Kay, J., Haucke, T., Stathatos, S., Deng, S., Young, E., Perona, P., Beery, S., & Van Horn, G. (2024). Align and distill: Unifying and improving domain adaptive object detection. arXiv preprint arXiv:2403.12029.

Kennerley, M., Wang, J.-G., Veeravalli, B., & Tan, R. T. (2024). Cat: Exploiting inter-class dynamics for domain adaptive object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16541–16550).

Liu, Y.-C., Ma, C.-Y., He, Z., Kuo, C.-W., Chen, K., Zhang, P., Wu, B., Kira, Z., & Vajda, P. (2021). Unbiased teacher for semi-supervised object detection. arXiv preprint arXiv:2102.09480.

Lv, J., Ren, Y., & Feng, G. (2025). Joint utilization of positive and negative pseudo-labels in semi-supervised facial expression recognition. *Pattern Recognition*, *159*, Article 111147.

Lv, F., Xia, C., Wang, S., & Cao, H. (2025). RT-DATR: Real-time unsupervised domain adaptive detection transformer with adversarial feature learning. (pp. arXiv–2504). ArXiv E-Prints.

Madi, B., Alaasam, R., Droby, A., & El-Sana, J. (2021). Advertisement extraction using deep learning. In *Document analysis and recognition–ICDAR 2021 workshops: Lausanne, Switzerland, September 5–10, 2021, proceedings, part II 16* (pp. 81–97). Springer.

Mattolin, G., Zanella, L., Ricci, E., & Wang, Y. (2023). Confmix: Unsupervised domain adaptation for object detection via confidence-based mixing. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 423–433).

Metro (2022–2023). Metro. Data extracted over the years 2022 and 2023. Accessed: 2023-04-26.

Nogueira-Rodríguez, A., Glez-Peña, D., Reboiro-Jato, M., & López-Fernández, H. (2023). Negative samples for improving object detection—A case study in AI-assisted colonoscopy for polyp detection. *Diagnostics*, *13*(5), 966.

Oza, P., Sindagi, V. A., Patel, V. M., et al. (2023). Unsupervised domain adaptation of object detectors: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(6), 4018–4040.

Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. arXiv preprint arXiv:2305.09972.

Rondan, N., Juayek, M., Joskowicz, J., Sotelo, R., Quincke, S., Patrone, A., Aguerre, M., & Gonzalez, G. (2025). Automatic detection of TV commercial breaks. In *2025 IEEE international conference on consumer electronics* (pp. 1–6). IEEE.

Sayyad, F. Z., Shallari, I., Mousavirad, S. J., O'Nils, M., & Qureshi, F. Z. (2025). AdVision: An efficient and effective deep learning based advertisement detector for printed media. *Machine Learning with Applications*, Article 100686.

Shehzadi, T., Hashmi, K. A., Sarode, S., Stricker, D., & Afzal, M. Z. (2025). STEP-DETR: Advancing DETR-based semi-supervised object detection with super teacher and pseudo-label guided text queries. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3069–3079).

Shehzadi, T., Hashmi, K. A., Stricker, D., & Afzal, M. Z. (2024). Sparse semi-detr: Sparse learnable queries for semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5840–5850).

Sohn, K., Zhang, Z., Li, C.-L., Zhang, H., Lee, C.-Y., & Pfister, T. (2020). A simple semi-supervised learning framework for object detection. (pp. arXiv–2005). ArXiv E-Prints.

Sun, T., Cui, C., Du, Y., & Liu, Y. (2025). PP-DocLayout: A unified document layout detection model to accelerate large-scale data construction. arXiv preprint arXiv:2503.17213.

Sydsvenskan (2022–2023). Sydsvenskan. Data extracted over the years 2022 and 2023. Accessed: 2023-04-26.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*(2), 99–114.

Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., & Le, X. (2022). Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4248–4257).

Wang, X., Yang, X., Zhang, S., Li, Y., Feng, L., Fang, S., Lyu, C., Chen, K., & Zhang, W. (2023). Consistent-teacher: Towards reducing inconsistent pseudo-targets in semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3240–3249).

Wang, S., Zhang, L., Chen, W., Wang, F., & Li, H. (2022). Refining pseudo labels for unsupervised domain adaptive re-identification. *Knowledge-Based Systems*, *242*, Article 108336.

Wardana, M. Z. P., & Wibowo, M. E. (2023). Audio-visual cnn using transfer learning for tv commercial break detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *17*(3), 291–300.

Wu, J., Li, Z., Sun, W., Yin, J., Nie, L., & Lin, Z. (2025). ClusMatch: Improving deep clustering by unified positive and negative pseudo-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., & Liu, Z. (2021). End-to-end semi-supervised object detection with soft teacher. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3060–3069).

Ye, K., & Kovashka, A. (2018). Advise: Symbolism and external knowledge for decoding advertisements. In *Proceedings of the European conference on computer vision* (pp. 837–855).

Yu, H., Li, Y., Wu, Q., Zhao, Z., Chen, D., Wang, D., & Wang, L. (2023). Mining negative temporal contexts for false positive suppression in real-time ultrasound lesion detection. In *International conference on medical image computing and computer-assisted intervention* (pp. 3–13). Springer.

Zakaryapour Sayyad, F., Shallari, I., Mousavirad, S. J., & O'Nils, M. (2024). Model evaluation and selection for robust and efficient advertisement detection in print media. In *International conference on advances in computing and data sciences* (pp. 211–224). Springer.

Zhao, Z., Kang, H., Wang, B., & He, C. (2024). DocLayout-YOLO: Enhancing document layout analysis through diverse synthetic data and global-to-local adaptive perception. (pp. arXiv–2410). ArXiv E-Prints.

Zheng, Y., Wu, J., Li, W., & Chen, Z. (2025). Universal domain adaptive object detection via dual probabilistic alignment. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 39* (pp. 10644–10652).

Zhou, H., Ge, Z., Liu, S., Mao, W., Li, Z., Yu, H., & Sun, J. (2022). Dense teacher: Dense pseudo-labels for semi-supervised object detection. In *European conference on computer vision* (pp. 35–50). Springer.

Zhu, C., Liu, K., Tang, W., Mei, K., Zou, J., & Huang, T. (2025). Hard-aware instance adaptive self-training for unsupervised cross-domain semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.