RESEARCH ARTICLE

Relaxed Bounding Boxes for Object Detection

Daniel Aioanei[1,*]

[1] Independent Scientist, 8400 Winterthur, Switzerland

Abstract

The Generalized Intersection over Union (GIoU) and the Manhattan distance between axis-aligned boxes represented either as corner coordinates or their center and size, are extended to accept a range of bounding boxes as ground truth, producing the metrics RIoU, R_1 **and** R_1^t**, respectively. In the context of Table Detection it is shown that this box relaxation procedure allows training object detection models with partial or inexact annotations. For the Table Structure Recognition task, several code improvements to Microsoft's open-source Table Transformer increase all** GriTS **metrics on PubTables-1M, with the overall accuracy increasing from** 0.8326 **to** 0.8433**. Then box relaxation is applied to take advantage in the object detection loss function of the discretizing nature of the post-inference table cell matrix extraction procedure. This further reduces the error of the** GriTS **metrics** Acc_{Con}**,** $GriTS_{Con}$**,** $GriTS_{Loc}$ **and** $GriTS_{Top}$ **on the PubTables-1M tables without spanning cells by** 1.8%**,** 13.2%**,** 10.6% **and** 14.9%**, respectively.**

Keywords: object detection, table detection, table structure recognition, bounding box regression, loss function.

Submitted: 08 August 2025
Accepted: 07 September 2025
Published: 17 September 2025

Vol. 1, **No.** 3, 2025.
10.62762/JIAP.2025.507329

***Corresponding author:**
✉ Daniel Aioanei
aioaneid@gmail.com

1 Introduction

With typical hard loss functions inconsistent or missing annotations have a detrimental effect on model performance [17, 22]. In object detection, missing annotations have previously been addressed through hard-example mining, by ignoring negatives that do not significantly overlap with positive instances [3] or with softer strategies such as reducing gradient magnitude as a function of overlap with positive examples [18]. Missing ground-truth bounding boxes have been the subject of Weakly-Supervised Object Detection (WSOD), where only image-level annotations are available [20], or of Semi-Supervised Object Detection (SSOD), which can also take into account clean bounding-box annotations [15].

Herein a generic approach is taken by constructing loss functions which treat all labels as exact but relax the ground-truth boxes so they represent a full range of boxes simultaneously, of which image-level annotations like those in WSOD, clean bounding-box annotations like in SSOD or noisy bounding-box priors are only special cases. The proposed method complements existing robust object detection algorithms, e.g. by allowing a strong prior [7] to be applied not only to one noisy bounding box, but to a full range of boxes which is likely to contain the ground truth.

As a practical demonstration these extended loss functions are integrated into the DEtection TRansformer (DETR) [2] and applied via the TAble

Citation
Aioanei, D. (2025). Relaxed Bounding Boxes for Object Detection. *ICCK Journal of Image Analysis and Processing*, 1(3), 107–124.

TRansformer (TATR) [11] to Table Detection (TD) and Table Structure Recognition (TSR). Specifically, it is shown that:

- A TD model can be successfully trained to detect multiple tables in images even when only one table per image is annotated with a bounding box;
- A small amount of box relaxation in TD has little impact on the COCO metrics [6], while significantly reducing the cardinality error;
- The TSR performance on tables without spanning cells can be improved by incorporating an approximation of the post-inference extraction step of the table cell matrix into the loss function, in the form of box relaxation.

For the TSR task, state-of-the-art Grid Table Similarity metrics (GriTS) [13] are achieved twice: first by improving the code of the open-source TATR implementation, and second by relaxing the ground-truth boxes under constraints imposed by the matrix cell-extraction step.

On the subset of *complex* tables of the PubTables-1M dataset, *i.e.* those containing spanning cells, a prevalent class of annotation errors is identified, opening the way for future improvements.

1.1 Paper Organization

The remainder of the paper is organized as follows:

- Section 2 surveys common loss and metric functions for object detection;
- Sections 3 to 5 generalize existing object detection loss functions to accommodate *relaxed* ground-truth boxes, each consisting of a range of boxes;
- Section 6 discusses how to incorporate non-injective post-inference steps as relaxed boxes during training, and how to apply the technique to TSR;
- Section 7 shows that a TD model can be trained by annotating the bounding box of only one table in each image, or by slightly relaxing the ground-truth boxes. Each of these approaches either preserves or reduces the COCO object cardinality error, respectively. Furthermore, box relaxation is shown to improve the TSR performance of TATR on simple tables;
- Section 8 highlights possible extensions of the method;
- Section 9 summarizes the results and concludes the paper.

2 Related Work

Generalized Intersection over Union (GIoU) [10] is a popular loss function and evaluation metric for object detection. It improves upon its scale-invariant predecessor, Intersection over Union (IoU), by addressing the gradient-vanishing problem in the non-overlapping case.

Many variations of (G)IoU have recently been developed, including:

- Bounded IoU (BIoU) considers the upper bound of IoU obtained when the three coordinates other than either the center X or Y coordinate, or the width or height of the predicted box, match the target box exactly [16];
- Distance-IoU (DIoU) incorporates the distance between the predicted and target box centers, normalized by the diagonal of the smallest enclosing box and squared [23];
- Complete IoU (CIoU) builds upon Distance-IoU (DIoU) by incorporating a suitably scaled squared difference between the arctangents of the aspect ratios of the predicted and target boxes [23];
- Gaussian Guided IoU (GGIoU) combines IoU with a Gaussian penalty that encourages the predicted and target box centers to be close [4];
- Alpha-IoU (α-IoU) applies a power transformation to IoU and related loss functions [5];
- Minimum Point Distance IoU (MPDIoU) uses the Euclidean distance between the top-left and bottom-right corners of the boxes [8];
- Corner-Point and Foreground-Area IoU (CFIoU) considers the distance between corresponding corner points. If the box centers coincide, it also considers the fraction of the minimum enclosing region covered by the target box; otherwise, it uses the area difference normalized by the minimum enclosing region [1];
- 3D-GIoU extends GIoU to three dimensions [21];
- Marginalized GIoU (MGIoU) is applied to convex shapes by computing the average one-dimensional GIoU across projections of both shapes onto the union of their normals [24].

A more traditional loss and metric function for bounding-box detection is the scale- and representation-dependent L_p loss ($0 \leq p \leq \infty$) applied to axis-aligned bounding boxes.

DETR, which uses the Hungarian matching algorithm as an alternative to anchor-based unordered set detection [19], relies on both GIoU and L_1 losses as follows:

- For each set of object predictions from the DETR decoder, an optimal bipartite matching with the ground-truth objects is computed. The pairwise matching cost is a linear function of:
  - L_1 loss between predicted and target boxes in center–size format,
  - GIoU between predicted and target boxes, and
  - the predicted probability of the target class.
- The final loss function is then computed as a linear function of the L_1 loss, GIoU, and class cross-entropy.

Both GIoU and the L_1 loss are generalized next so that the ground truth consists of a range of boxes bounded by an optional hole border [14] and/or an optional outer border.

3 Relaxed Intersection over Union (RIoU)

Definition 1 (RIoU: Relaxed Intersection over Union)**.** *Let $\mathcal{C}$ be one of the following sets: the full set of closed convex sets in $\mathbb{R}^d$, the set of d-dimensional boxes, or the set of d-dimensional axis-aligned boxes. Given*

- *a* predicted shape $B \in \mathcal{C}$,
- *an optional shape $H \in \mathcal{C}$, referred to as the* hole border, *and*
- *an optional shape $O \in \mathcal{C}$, referred to as the* outer border,

let

$$\begin{aligned}\mathrm{RIoU}(B,H,O) = \frac{|B\cap H|}{|H|}\cdot\frac{|O|}{|O\cup B|} - 1 \\ + \frac{|B\cup H|}{|E(B,H)|}/2 + \frac{|O\cup B|}{|E(O,B)|}/2\end{aligned} \tag{1}$$

where $E(P,Q)$ denotes the convex hull in $\mathcal{C}$ of shapes $P, Q \in \mathcal{C}$ and $|\cdot|$ denotes the volume of a shape.

The definition of RIoU is extended to empty or unspecified borders according to the following principles:

- An unspecified hole border is treated as $H \subseteq B$;
- An unspecified outer border is treated as $B \subseteq O$;
- If hole border H is specified, the relation $H \subseteq B$ is required to maximize $\mathrm{RIoU}(B, H, \cdot)$;
- If outer border O is specified, the relation $B \subseteq O$ is required to maximize $\mathrm{RIoU}(B, \cdot, O)$;

RIoU is consequently endowed with the following conventions:

- If H is unspecified then $|B \cap H|/|H| = |B \cup H|/|E(H,B)| = 1$;
- If O is unspecified then $|O|/|O \cup B| = |O \cup B|/|E(O,B)| = 1$;
- If $|H| = 0$ then $|B \cap H|/|H| = 1$ if $H \subseteq B$ and 0 otherwise;
- If $|O \cup B| = 0$, then $|O|/|O \cup B| = 1$ if $B \subseteq O$ and 0 otherwise;
- If $|E(P,Q)| = 0$, then $|P \cup Q|/E(P,Q) = 1$ if $P \subseteq Q$ or $Q \subseteq P$ and 0 otherwise.

A few properties follow immediately:

- It always holds that $-1 \leq \mathrm{RIoU} \leq 1$;
- If $H = O$, then RIoU equals GIoU;
- With both borders specified, $\mathrm{RIoU}(B,H,O) = 1 \iff H \subseteq B \subseteq O$;
- With only H specified, $-½ \leq \mathrm{RIoU}(B,H,\_) \leq 1$;
- With only O specified, $-½ \leq \mathrm{RIoU}(B,\_,O) \leq 1$;
- With both borders unspecified, $\mathrm{RIoU}(B,\_,\_) = 1$.

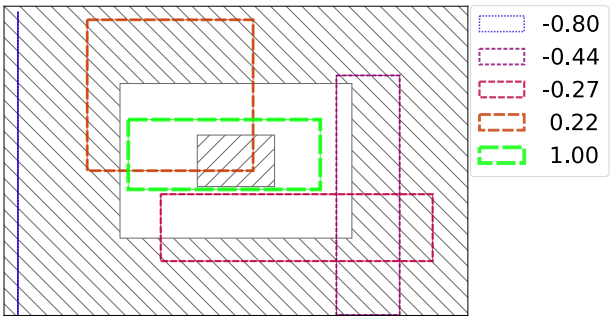
Figure 1 shows examples of RIoU in the common case where $\mathcal{C}$ is the set of axis-aligned rectangles in 2D.

4 Relaxed L_p Distance Between Axis-Aligned Boxes (R_p)

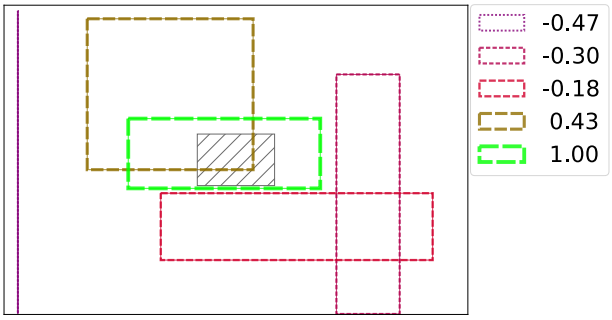
Let an axis-aligned box in $\mathbb{R}^d$ be represented by the coordinates of its lower-bound corner $l = l_1, \ldots, l_d$ and its upper-bound corner $u = u_1, \ldots, u_d$, where $l_i \leq u_i, i = 1..d$. An axis-aligned box can be *partially specified*, *i.e.* any subset of the coordinates of l or u may be left unspecified.

Definition 2 (R_p: Relaxed L_p distance of axis-aligned boxes)**.** *Given*

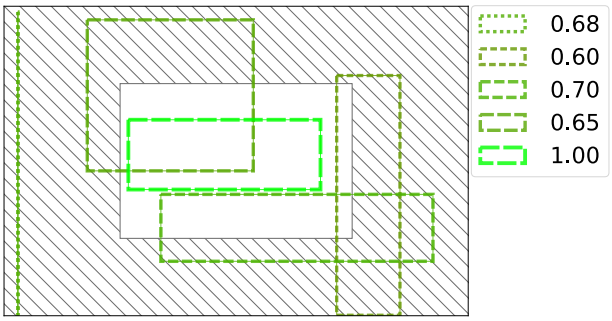
- *an arbitrary axis-aligned box $B = (l^b, u^b)$, referred to as the* predicted box,



(a) $RIoU(B, H, O)$ for rectangles B relative to hole and outer borders $H \subseteq O$. The maximum value of 1 is reached if and only if $H \subseteq B \subseteq O$.



(b) $RIoU(B, H, \_)$ for rectangles B relative to hole border H. The maximum value of 1 is reached iff $H \subseteq B$.



(c) $RIoU(B, H, \_)$ for rectangles B relative to outer border O. The maximum value of 1 is reached if and only if $B \subseteq O$.

Figure 1. $RIoU(B, \cdot, \cdot)$ for axis-aligned boxes in 2D. Penalty areas are hatched. Legends display numerical $RIoU$ values and the rectangle color palette follows a piecewise-linear scale from blue (-1) through red (0) to green (1).

- *a partially specified axis-aligned box* $H = (l^h, u^h)$, *referred to as the* hole border, *and*
- *a partially specified axis-aligned box* $O = (l^o, u^o)$, *referred to as the* outer border,

let

$$R_p(B, H, O) = \min_{H \subseteq J \subseteq O} L_p(B, J), \tag{2}$$

where $0 \leq p \leq \infty$, L_p *is the distance induced by the p-norm in the* $\mathbb{R}^{2d}$ *Cartesian coordinate space, and* $J = (l^j, u^j)$ *represents any axis-aligned box which includes* H *and is itself included in* O, i.e. $l_i^o \leq l_i^j \leq l_i^h$ *and* $u_i^h \leq u_i^j \leq u_i^o$ *for* $i = 1..d$.

The axis-aligned boxes J in Definition 2 are said to be *compatible* with the hole border H and the outer border O.

By convention, whenever l_i^o, l_i^h, u_i^h, or u_i^o are unspecified in Definition 2 the corresponding inequalities are removed. Moreover, if $H = O$, then R_p reduces to the standard L_p distance between axis-aligned boxes represented by their lower-bound and upper-bound corners.

The computation of R_p is given in Algorithm 1.

Algorithm 1: R_p: Relaxed L_p distance of axis-aligned boxes

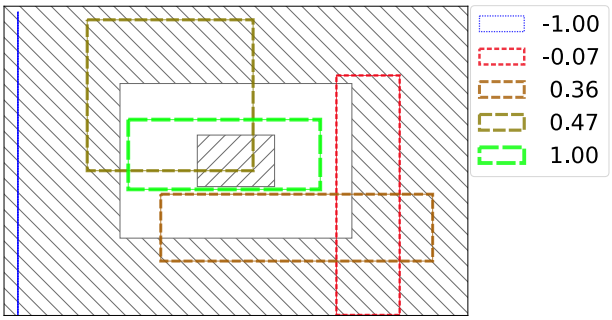
Input: $B = (l^b, u^b)$, $H = (l^h, u^h)$, $O = (l^o, u^o)$
for $i = 1$ **to** d **do**
  if l_i^o is unspecified **then**
    $l_i^o \leftarrow -\infty$
  end if
  if u_i^o is unspecified **then**
    $u_i^o \leftarrow \infty$
  end if
  if l_i^h is unspecified **then**
    $l_i^h \leftarrow u_i^o$
  end if
  if u_i^h is unspecified **then**
    $u_i^h \leftarrow l_i^o$
  end if
  $l_i^x \leftarrow \min(\max(l_i^b, l_i^o), l_i^h)$ {Clamp l_i^b }
  $u_i^x \leftarrow \min(\max(u_i^b, u_i^h), u_i^o)$ {Clamp u_i^b }
end for
$X \leftarrow (l^x, u^x)$ {Nearest compatible neighbor}
return $||B - X||_p$ {Optimal distance}

Proof of Algorithm 1. The correctness follows because each dimension can be clamped in independently. Thus, the optimal compatible box can be constructed dimension by dimension. □

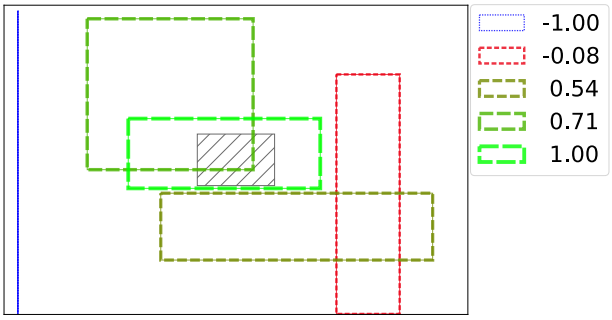
Some 2D examples of R_1 are shown in Figure 2.

5 Relaxed L_1 Distance in Center-Size Format (R_1^t)

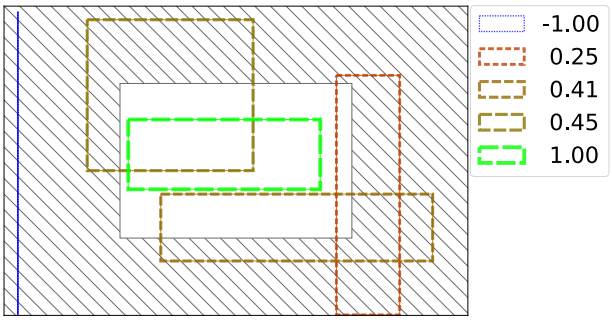
In DETR, the distance between axis-aligned boxes is not computed directly from the lower- and upper-bound corner coordinates. Instead, it is measured between their box centers and sizes.



(a) Expression $1 - 2R_1(B, H, O)/\max_B R_1$ for rectangles B relative to hole and outer borders $H \subseteq O$. R_1 reaches its minimum value 0 if and only if $H \subseteq B \subseteq O$.



(b) Expression $1 - 2R_1(B, H, \_)/\max_B R_1$ for rectangles B relative to hole border H. R_1 reaches its minimum value 0 if and only if $H \subseteq B$.



(c) Expression $1 - 2R_1(B, \_, O)/\max_B R_1$ for rectangles B relative to outer border O. R_1 reaches its minimum value 0 if and only if $B \subseteq O$.

Figure 2. Expression $1 - 2R_1(B, \cdot, \cdot)/\max_B R_1(B, \cdot, \cdot)$ for axis-aligned boxes in 2D. Penalty areas are hatched. The color palette and the boxes B, hole borders H and outer borders O are as in Figure 1.

A simple two-step algorithm can be used to find the nearest compatible L_1 neighbor in the center–size format. For each dimension:

1. Clamp the *size* of the interval between the minimum and maximum size of any box compatible with the target relaxed box. The center location does not affect this step;
2. Move the *center* of the interval by the smallest amount necessary – while keeping its size fixed as determined in the previous step – so that it becomes compatible with the target relaxed box.

A formal treatment of a broader class of linearly transformed boxes is provided in Appendix B, of which the center–size representation is a special case which is illustrated in Appendix C.

6 Training for Post-Inference Processing

Object detection is often followed by post-processing steps exhibiting constant regions, e.g. by partially discretizing the predicted bounding boxes. This effectively creates equivalence classes of image annotations.

If a subset of the ground-truth boxes in a training example can be relaxed simultaneously without any compatible combination of boxes leaving the equivalence class, then the training objective can be adapted to target not only the original example, but a larger subset of its equivalence class. This procedure can improve performance by providing better alignment of the training objective, while also offering several practical advantages:

- Box relaxation within the equivalence class only needs to be applied to ground-truth examples. Since these are already curated, their post-processing can be simpler than that of predicted boxes.
- Ground-truth boxes for which the objective function is not demonstrably constant in a local region can still be included in the training set, albeit without relaxation.
- There is no need to design a new loss or bipartite matching cost function for every post-processing algorithm. The same loss functions developed previously can be applied without modification.

6.1 TSR with Relaxed Boxes under GriTS Equivalence

It is next shown how the box relaxation procedure can be applied to table structure recognition (TSR).

A set of metrics sharing a common framework has been proposed to evaluate the performance of table cell content, location, and topology recognition: $GriTS_{Con}$, $GriTS_{Loc}$, and $GriTS_{Top}$. The metric Acc_{Con} has also been derived as the case where $GriTS_{Con} = 1$.

$GriTS_{Con}$ relies for cell-partial correctness on the normalized length of the longest contiguous junk-free matching subsequence. $GriTS_{Loc}$ and $GriTS_{Top}$ rely instead on the intersection area divided by the area of

the minimal box enclosing both predicted and target boxes (according to the open-source implementation, which differs from the "IoU" designation in [13]).

These metrics rely on an approximate matrix similarity algorithm applied to table cell matrices, which represent strongly discretized versions of the target and predicted boxes.

Box relaxation under table cell matrix equivalence is formulated as a constrained optimization problem in which the hole and outer borders serve as optimization variables. The optimization objective is defined as the sum of the differences between the perimeters of the outer and hole borders. The constraints which ensure table cell matrix equivalence fall into three categories:

- Bound constraints, e.g. the hole border of each object must contain the center of the original box;
- Linear constraints, e.g. a row underlying a header must have substantial vertical overlap with the header;
- Non-linear constraints, e.g. on the amount of area overlap between spanning cells and simple cells.

The resulting relaxed boxes for a PubTables-1M training example is shown in Figure 3.

Additional details of the constrained optimization problem are provided in Appendix D.

7 Experimental Results

The experiments were performed on an NVIDIA GeForce RTX 3070 with an Intel i9-9900K CPU and a Gen-4 ThinkPad Laptop. The source code is publicly available at https://github.com/aioaneid/table-transformer.

7.1 Table Detection (TD) with Box Relaxation

The TATR model for TD was trained using the following procedures:

[A] Using the full PubTables-1M training dataset for table detection;

[B] Restricting the training data to images containing only a single table;

[C] Including all images, but in multi-table images keeping only one randomly selected table;

[D] Including all images, but in multi-table images keeping the bounding box of a single randomly selected table, while relaxing the bounding boxes of the others so their outer borders cover the full image and hole borders are absent. This is equivalent to counting objects by category, *i.e.* either *table* or *table rotated*;

[E] Including all images, and relaxing each bounding box by shrinking and expanding each dimension by 2 pixels around the box center to obtain the hole and outer borders, respectively. The tight-annotation crop of TATR was configured to crop around the outer border;

[F] Including all images, and symmetrically relaxing each bounding box in both dimensions by 4 pixels, or less if required to maintain symmetry, while keeping the box center fixed. The tight-annotation crop of TATR was configured to crop around the midline between the hole and outer borders, which coincides with the original bounding box;

[G] Same as Model [F], but with 8 pixels of relaxation instead of 4.

As shown in Table 1, the cardinality error – measured as the average absolute error in the number of predicted objects – is highest when only one table is annotated in multi-table images (Models [B] and [C]). With box relaxation, however, the remaining tables can be counted by object category (*table* or *table rotated*) (Model [D]), restoring the cardinality error to the level observed with standard training (Model [A]).

A small amount of relaxation either has negligible impact on the COCO metrics (Model [E]), or slightly reduces them while also reducing the object cardinality error (Models [F] and [G]). In these experiments, the cardinality error consistently decreased with increasing box relaxation.

Table 1. The best – across all 20 training epochs – Cardinality Error, COCO Average Precision (AP) and COCO Average Recall (AR) for each TD model described in Section 7.1. In each column, the best and worst results are shown in **bold** and *italic*, respectively.

Model	Card. Error	AP	AR	Training Dataset
Model [A]	0.0018	**0.9800**	**0.9900**	Full PubTables-1M training dataset
Model [B]	*0.1050*	*0.8700*	*0.8870*	Only images containing a single table
Model [C]	0.0186	0.9770	0.9880	All images, but in each image keeping 1 randomly selected table
Model [D]	0.0018	0.9730	0.9880	All images, 1 random table/image and counting the other tables
Model [E]	0.0018	**0.9800**	**0.9900**	All images with all tables shrunk and expanded by 2 pixels
Model [F]	0.0016	0.9790	**0.9900**	All images with all tables shrunk and expanded by 4 pixels
Model [G]	**0.0014**	0.9760	0.9850	All images with all tables shrunk and expanded by 8 pixels

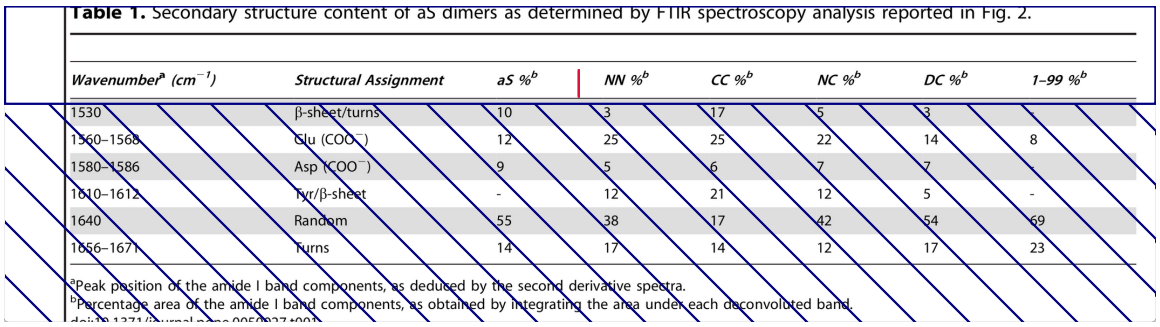


Table 1. Secondary structure content of aS dimers as determined by FTIR spectroscopy analysis reported in Fig. 2.

Wavenumber[a] (cm^{-1})	*Structural Assignment*	*aS %[b]*	*NN %[b]*	*CC %[b]*	*NC %[b]*	*DC %[b]*	*1–99 %[b]*
1530	β-sheet/turns	10	3	17	5	3	-
1560–1568	Glu (COO^-)	12	25	25	22	14	8
1580–1586	Asp (COO^-)	9	5	6	7	7	-
1610–1612	Tyr/β-sheet	-	12	21	12	5	-
1640	Random	55	38	17	42	54	69
1656–1671	Turns	14	17	14	12	17	23

[a]Peak position of the amide I band components, as deduced by the second derivative spectra.
[b]Percentage area of the amide I band components, as obtained by integrating the area under each deconvoluted band.

(a) Relaxed column header. The hole border is a short vertical segment.

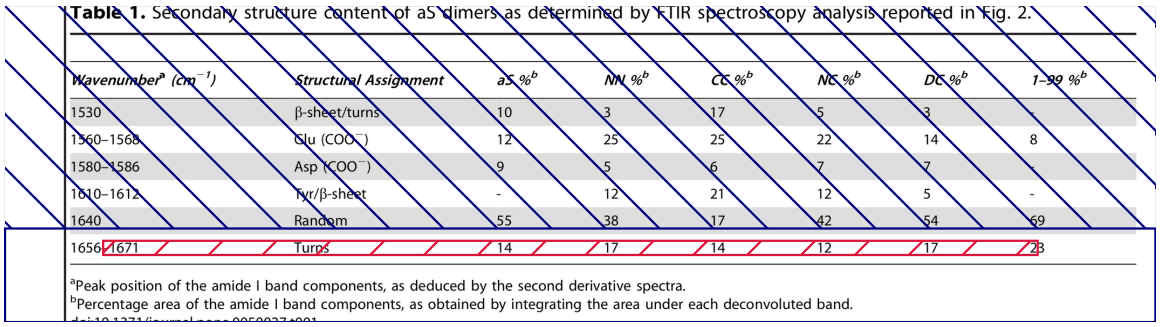


Table 1. Secondary structure content of aS dimers as determined by FTIR spectroscopy analysis reported in Fig. 2.

Wavenumber[a] (cm^{-1})	*Structural Assignment*	*aS %[b]*	*NN %[b]*	*CC %[b]*	*NC %[b]*	*DC %[b]*	*1–99 %[b]*
1530	β-sheet/turns	10	3	17	5	3	-
1560–1568	Glu (COO^-)	12	25	25	22	14	8
1580–1586	Asp (COO^-)	9	5	6	7	7	-
1610–1612	Tyr/β-sheet	-	12	21	12	5	-
1640	Random	55	38	17	42	54	69
1656–1671	Turns	14	17	14	12	17	23

[a]Peak position of the amide I band components, as deduced by the second derivative spectra.
[b]Percentage area of the amide I band components, as obtained by integrating the area under each deconvoluted band.

(b) Relaxed last row, which can extend indefinitely downward.

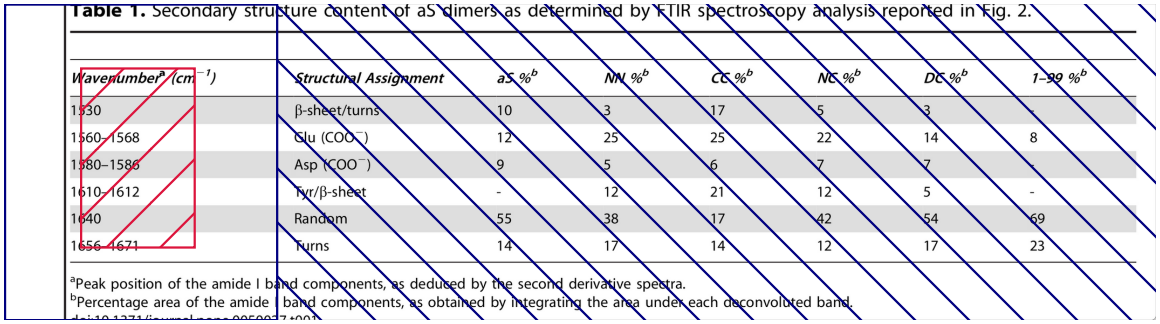


Table 1. Secondary structure content of aS dimers as determined by FTIR spectroscopy analysis reported in Fig. 2.

Wavenumber[a] (cm^{-1})	*Structural Assignment*	*aS %[b]*	*NN %[b]*	*CC %[b]*	*NC %[b]*	*DC %[b]*	*1–99 %[b]*
1530	β-sheet/turns	10	3	17	5	3	-
1560–1568	Glu (COO^-)	12	25	25	22	14	8
1580–1586	Asp (COO^-)	9	5	6	7	7	-
1610–1612	Tyr/β-sheet	-	12	21	12	5	-
1640	Random	55	38	17	42	54	69
1656–1671	Turns	14	17	14	12	17	23

[a]Peak position of the amide I band components, as deduced by the second derivative spectra.
[b]Percentage area of the amide I band components, as obtained by integrating the area under each deconvoluted band.

(c) Relaxed first column, which can extend indefinitely to the left.

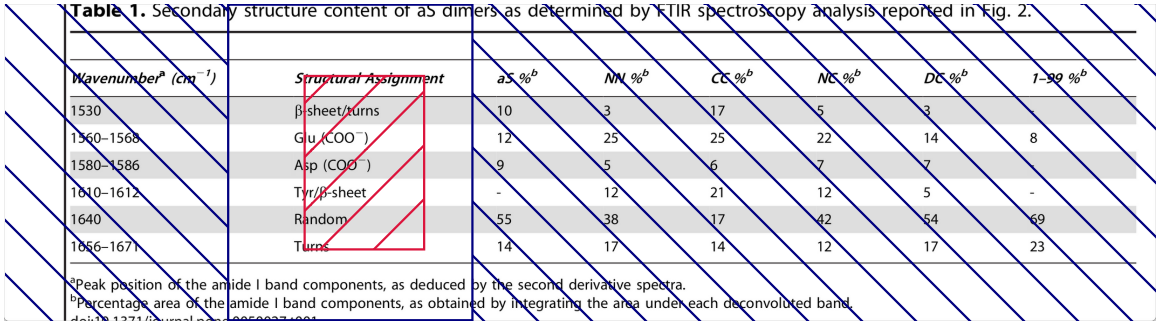


Table 1. Secondary structure content of aS dimers as determined by FTIR spectroscopy analysis reported in Fig. 2.

Wavenumber[a] (cm^{-1})	*Structural Assignment*	*aS %[b]*	*NN %[b]*	*CC %[b]*	*NC %[b]*	*DC %[b]*	*1–99 %[b]*
1530	β-sheet/turns	10	3	17	5	3	-
1560–1568	Glu (COO^-)	12	25	25	22	14	8
1580–1586	Asp (COO^-)	9	5	6	7	7	-
1610–1612	Tyr/β-sheet	-	12	21	12	5	-
1640	Random	55	38	17	42	54	69
1656–1671	Turns	14	17	14	12	17	23

[a]Peak position of the amide I band components, as deduced by the second derivative spectra.
[b]Percentage area of the amide I band components, as obtained by integrating the area under each deconvoluted band.

(d) Relaxed second column, allowing some overlap with nearby columns.

Figure 3. Subset of hole and outer border pairs for a training TSR image [9]. Hole border penalty areas are diagonally hatched in crimson, while outer border penalty areas are back-diagonally hatched in navy blue.

Figure 4 shows a validation image where the standard model (Model [A]) detects a spurious table, while the model trained with 8 pixels of box relaxation (Model [G]) correctly identifies all tables.

Another validation example, shown in Figure 5, illustrates a case where both the standard model (Model [A]) and the 8-pixel relaxation model (Model [G]) detect a spurious table, but at different

SRS Dataset	Average AUC (MGPS)	Average AUC (MCEM MGPS)
All ADRs	0.6787	**0.7225**
Acute Myocardial Infarction (AMI)	0.5834	**0.6109**
Acute Liver Injury(ALI)	**0.6659**	0.6512
Acute Renal Failure (ARF)	0.6926	**0.8243**
Upper GI Bleeding (UGB)	0.6610	**0.7660**

Table 2. Comparison of the standard MGPS score and MCEM MGPS score based on ADRs of interest in FAERS.

SRS Dataset	Average AUC (MGPS)	Average AUC (MCEM MGPS)
All ADRs	0.7366	**0.7683**
Acute Myocardial Infarction (AMI)	0.7500	**0.7812**
Acute Liver Injury(ALI)	**0.6829**	0.4756
Upper GI Bleeding (UGB)	0.7500	**0.7870**

Table 3. Comparison of the standard MGPS score and MCEM MGPS score based on ADRs of interest in MedEffect.

SRS Dataset	AUC (MGPS)	AUC (MCEM MGPS)
As of FAERS 2007	0.6994	**0.7265**
As of FAERS 2008	0.6936	**0.7067**
As of FAERS 2009	0.6889	**0.7096**
As of FAERS 2010	0.6853	**0.7282**
As of FAERS 2011	0.6868	**0.7353**
As of FAERS 2012	0.6907	**0.7146**
As of FAERS 2013	0.7091	**0.7261**
As of FAERS 2014	0.7012	**0.7385**

Table 4. Comparison of the standard MGPS score and MCEM MGPS score by reporting ending years.

(a) Table detection by the standard model (Model [A]).

SRS Dataset	Average AUC (MGPS)	Average AUC (MCEM MGPS)
All ADRs	0.6787	**0.7225**
Acute Myocardial Infarction (AMI)	0.5834	**0.6109**
Acute Liver Injury(ALI)	**0.6659**	0.6512
Acute Renal Failure (ARF)	0.6926	**0.8243**
Upper GI Bleeding (UGB)	0.6610	**0.7660**

Table 2. Comparison of the standard MGPS score and MCEM MGPS score based on ADRs of interest in FAERS.

SRS Dataset	Average AUC (MGPS)	Average AUC (MCEM MGPS)
All ADRs	0.7366	**0.7683**
Acute Myocardial Infarction (AMI)	0.7500	**0.7812**
Acute Liver Injury(ALI)	**0.6829**	0.4756
Upper GI Bleeding (UGB)	0.7500	**0.7870**

Table 3. Comparison of the standard MGPS score and MCEM MGPS score based on ADRs of interest in MedEffect.

SRS Dataset	AUC (MGPS)	AUC (MCEM MGPS)
As of FAERS 2007	0.6994	**0.7265**
As of FAERS 2008	0.6936	**0.7067**
As of FAERS 2009	0.6889	**0.7096**
As of FAERS 2010	0.6853	**0.7282**
As of FAERS 2011	0.6868	**0.7353**
As of FAERS 2012	0.6907	**0.7146**
As of FAERS 2013	0.7091	**0.7261**
As of FAERS 2014	0.7012	**0.7385**

Table 4. Comparison of the standard MGPS score and MCEM MGPS score by reporting ending years.

(b) Table detection by the model trained with 8 pixels of box relaxation (Model [G]).

Figure 4. A validation image excerpt in which the standard model [25] (Model [A]) finds a spurious table (a), while the model trained with 8 pixels of box relaxation (Model [G]) correctly identifies all tables (b).

locations.

These results confirm that box relaxation enables the training of object detection models using partial or relaxed annotations, without introducing performance-damaging contradictions in the training dataset.

Table 1. The *p*-value of the percentage of viable mouse glioma cell line CT2A cells after exposure to a 100 µT electromagnetic field (EMF) at 20 Hz.

20 Hz	*p* Value
24 h	0.435
48 h	0.005
72 h	0.207

Table 2. The *p*-value of the percentage of viable CT2A cells after exposure to a 100 µT EMF at 30 Hz.

30 Hz	*p* Value
24 h	0.007
48 h	0.002
72 h	0.001

Table 3. The *p*-value of the percentage of viable CT2A cells after the exposure to a 100 µT EMF at 50 Hz.

50 Hz	*p* Value
24 h	0.009
48 h	0.299
72 h	0.002

(a) The standard model (Model [A]) duplicates the first table.

Table 1. The *p*-value of the percentage of viable mouse glioma cell line CT2A cells after exposure to a 100 µT electromagnetic field (EMF) at 20 Hz.

20 Hz	*p* Value
24 h	0.435
48 h	0.005
72 h	0.207

Table 2. The *p*-value of the percentage of viable CT2A cells after exposure to a 100 µT EMF at 30 Hz.

30 Hz	*p* Value
24 h	0.007
48 h	0.002
72 h	0.001

Table 3. The *p*-value of the percentage of viable CT2A cells after the exposure to a 100 µT EMF at 50 Hz.

50 Hz	*p* Value
24 h	0.009
48 h	0.299
72 h	0.002

(b) The model trained with 8 pixels of box relaxation (Model [G]) duplicates the second table.

Figure 5. Excerpt from a validation image showing TD inferences of the standard model and the model trained with 8-pixel box relaxation [26].

7.2 Table Structure Recognition (TSR) with Box Relaxation

A stronger baseline than TATR v1.1 was established first by fixing the following issues in the open-source implementation:

- In some cases, disjoint boxes were incorrectly considered to have a non-zero intersection area;
- The row and column alignment code unintentionally altered shared data structures;
- Multiple overlapping predicted headers were incorrectly aligned;
- The state of the random number generator was lost when training was resumed.

As shown in Table 2, the resulting baseline significantly outperforms previously reported metrics.

As an object-detection model, the TATR TSR model benefits from the same advantages described in

Section 7.1. Training with missing annotations is feasible provided that objects are counted according to TSR categories: *table*, *table column*, *table row*, *table column header*, *table projected row header*, and *table spanning cell*. Partial annotations are also supported. For example, a zero-area table column hole border can be created with a single click inside a column (not shown).

Box relaxation can further improve TSR performance via the constrained box relaxation technique described in Section 6.1. In order to not interfere with TATR v1.1's tight annotation cropping, the table outer border was fixed to the original bounding box. Training with constrained box relaxation led to substantial improvements in the GriTS metrics for the category of simple tables, as shown in Table 2.

Table 2. GriTS performance of TATR v1.0, v1.1, v1.1 with bug fixes, and v1.1 with bug fixes plus constrained table relaxation. In each row, the best result within its category is shown in **bold**. One epoch corresponds to all 758,849 PubTables-1M TSR training images. For comparison, the original TATR v1.1 model was previously trained for 30 epochs of 720,000 images each.

	All tables			
Metric	TATR v1.0	TATR v1.1	Bug fixes	Box relaxation
Acc_{Con}	0.8243	0.8326	0.8433	**0.8458**
$GriTS_{Con}$	0.9850	0.9855	0.9862	**0.9866**
$GriTS_{Loc}$	0.9786	0.9797	0.9806	**0.9811**
$GriTS_{Top}$	0.9849	0.9851	0.9858	**0.9861**
Epochs	20	28.5	28	28
	Simple tables (no spanning cells)			
Metric	TATR v1.1	Bug fixes	Box relaxation	
Acc_{Con}	0.9551	0.9661	**0.9667**	
$GriTS_{Con}$	0.9936	0.9947	**0.9954**	
$GriTS_{Loc}$	0.9922	0.9934	**0.9941**	
$GriTS_{Top}$	0.9943	0.9953	**0.9960**	
Epochs	28.5	28	28	
	Complex tables (with spanning cells)			
Metric	TATR v1.1	Bug fixes	Box relaxation	
Acc_{Con}	0.7186	0.7324	**0.7363**	
$GriTS_{Con}$	0.9777	0.9786	**0.9789**	
$GriTS_{Loc}$	0.9680	0.9693	**0.9697**	
$GriTS_{Top}$	0.9761	**0.9774**	0.9773	
Epochs	28.5	28	28	

A comparison on a simple image between the baseline model and the constrained box relaxation model is shown in Figure 6.

While constrained box relaxation consistently improved performance across epochs and GriTS metrics for simple tables, no clear pattern emerged for complex tables (Figure 10). Investigation revealed that approximately 1.4% of the complex tables contain a spanning cell deemed invalid by the cell-matrix extraction procedure.

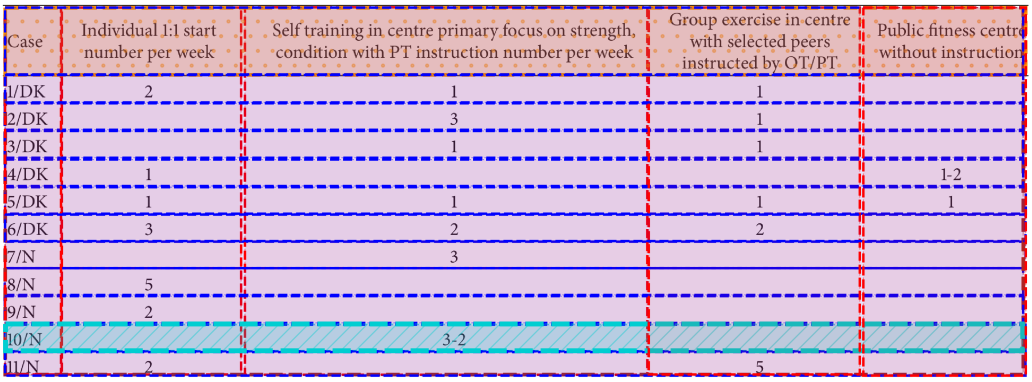
Such a validation example is shown in Figure 7. In the PubTables-1M annotation, the spanning cell at the beginning of the second row (Figure 7(a)) violates TATR's assumption of a tree structure of spanning cells within a header, causing it to be treated as invalid and dropped. Consequently, the text "Analysis without 12-gene risk score" is split into two simple cells (Figure 7(b)). The header is also incorrectly annotated as three rows instead of one, and a spurious vertical spanning cell encompasses the text "(0.84-2.16)". The baseline model (TATR v1.1 with bug fixes) also made multiple errors (Figure 8) but matched the PubTables-1M annotation somewhat closely, achieving $GriTS_{Con} = 0.9821$, $GriTS_{Loc} = 0.9770$, and $GriTS_{Top} = 0.9821$ on this image alone. By contrast, the constrained box relaxation model made a single mistake by missing a spanning cell (Figure 9), which by comparison to the PubTables-1M annotation results in lower metric values of $GriTS_{Con} = 0.9059$, $GriTS_{Loc} = 0.7048$ and $GriTS_{Top} = 0.8815$. This example highlights the need for curated complex-table annotations in PubTables-1M for accurate benchmarking.

To enable comparison with previously published results, PubTables-1M tables containing provably invalid spanning cells were not excluded. Given that these PubTables-1M annotations violate some of the constraints in Equation (6), they represent infeasible solutions to the numerical optimization problem. For this reason they were included in the training dataset without any relaxation.

7.3 Model Training Performance

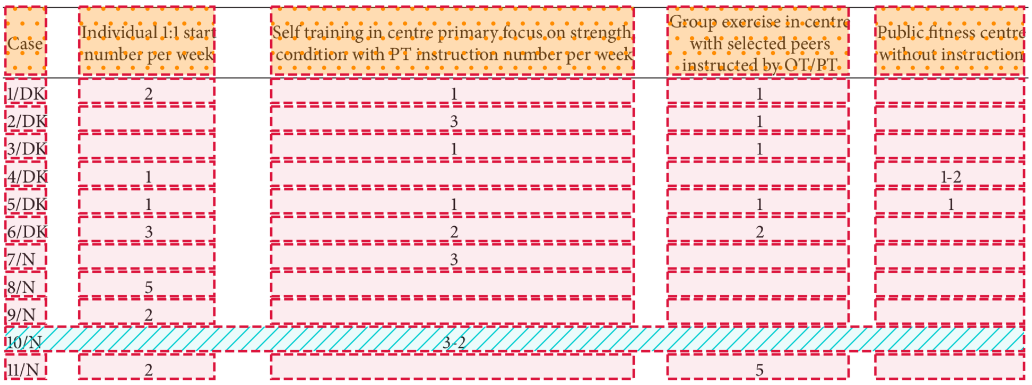
Since the loss function is modified only slightly and the model size remains unchanged, box relaxation has a marginal impact on training time per epoch. Indeed, in the experiments presented here, box relaxation increased the training time by about 1%. The inference procedure, on the other hand, remains unaffected.

The preparation of the training dataset with relaxed boxes must also be considered. This step is easily parallelizable, as there are no cross-image dependencies. While faster and more sophisticated numerical optimization tools exist, the TSR results reported here were obtained using a Python implementation of Algorithm 3, which takes approximately 10 seconds per image on a single laptop CPU core.



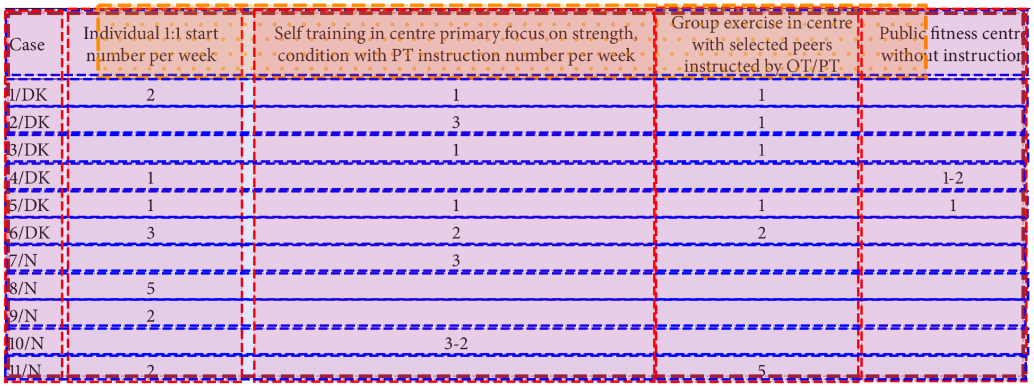
Case	Individual 1:1 start number per week	Self training in centre primary focus on strength, condition with PT instruction number per week	Group exercise in centre with selected peers instructed by OT/PT	Public fitness centre without instruction
1/DK	2	1	1	
2/DK		3	1	
3/DK		1	1	
4/DK	1			1-2
5/DK	1	1	1	1
6/DK	3	2	2	
7/N		3		
8/N	5			
9/N	2			
10/N		3-2		
11/N	2		5	

(a) The baseline model detects a spurious table projected row header.



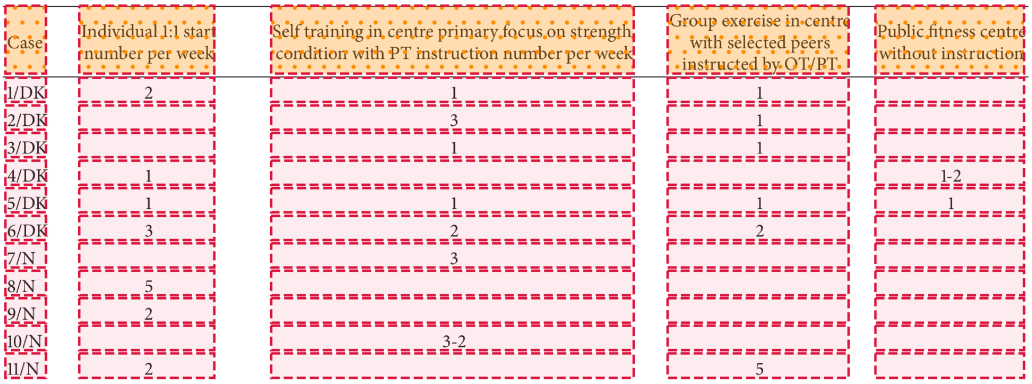
Case	Individual 1:1 start number per week	Self training in centre primary focus on strength, condition with PT instruction number per week	Group exercise in centre with selected peers instructed by OT/PT	Public fitness centre without instruction
1/DK	2	1	1	
2/DK		3	1	
3/DK		1	1	
4/DK	1			1-2
5/DK	1	1	1	1
6/DK	3	2	2	
7/N		3		
8/N	5			
9/N	2			
10/N		3-2		
11/N	2		5	

(b) The table cell matrix of the baseline model corresponding to (a).



Case	Individual 1:1 start number per week	Self training in centre primary focus on strength, condition with PT instruction number per week	Group exercise in centre with selected peers instructed by OT/PT	Public fitness centre without instruction
1/DK	2	1	1	
2/DK		3	1	
3/DK		1	1	
4/DK	1			1-2
5/DK	1	1	1	1
6/DK	3	2	2	
7/N		3		
8/N	5			
9/N	2			
10/N		3-2		
11/N	2		5	

(c) The model trained on relaxed boxes does not find any spurious objects.



Case	Individual 1:1 start number per week	Self training in centre primary focus on strength, condition with PT instruction number per week	Group exercise in centre with selected peers instructed by OT/PT	Public fitness centre without instruction
1/DK	2	1	1	
2/DK		3	1	
3/DK		1	1	
4/DK	1			1-2
5/DK	1	1	1	1
6/DK	3	2	2	
7/N		3		
8/N	5			
9/N	2			
10/N		3-2		
11/N	2		5	

(d) The table cell matrix of the model trained on relaxed boxes corresponding to (c), without errors.

Figure 6. Comparison of baseline and model trained with relaxed boxes on a simple table from PubTables-1M [27].

Variable*	*P*-value	Hazard Ratio (95% CI)	
Analysis without 12-gene risk score			
Gender (Male)	0.22	1.34	(0.84–2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95–2.74)
Tumor Stage			
Stage II	6.25E-05	2.91	(1.72–4.91)
Stage III	1.09E-05	4.16	(2.20–7.85)
Analysis with 12-gene risk score			
Gender (Male)	0.17	1.40	(0.87–2.26)
Age at diagnosis (>60)	0.29	1.34	(0.78–2.31)
Tumor Stage			
Stage II	3.47E-04	2.61	(1.54–4.43)
Stage III	7.40E-06	4.31	(2.28–8.16)
12-gene risk score	**3.10E-05**	**3.94**	**(2.07–7.52)**

(a) The box annotations as they appear in the PubTables-1M validation dataset. The header, which in reality consists of only the top row, is incorrectly annotated in PubTables-1M as spanning the top three rows. Of the three spanning cell annotations (green, back diagonal hatch pattern), only the top-right one is correct.

Variable*	*P*-value	Hazard Ratio (95% CI)	
Analysis without 12-gene risk score			
Gender (Male)	0.22	1.34	(0.84–2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95–2.74)
Tumor Stage			
Stage II	6.25E-05	2.91	(1.72–4.91)
Stage III	1.09E-05	4.16	(2.20–7.85)
Analysis with 12-gene risk score			
Gender (Male)	0.17	1.40	(0.87–2.26)
Age at diagnosis (>60)	0.29	1.34	(0.78–2.31)
Tumor Stage			
Stage II	3.47E-04	2.61	(1.54–4.43)
Stage III	7.40E-06	4.31	(2.28–8.16)
12-gene risk score	**3.10E-05**	**3.94**	**(2.07–7.52)**

(b) The table cell matrix corresponding to the PubTables-1M box annotations of (a). The spanning cell holding together the text in the second row is deemed invalid by the TATR cell-matrix extraction procedure, so the row's text gets split in two cells.

Figure 7. A table image with an incorrect annotation in the PubTables-1M validation set [28]. The column header is colored dark orange with a dot hatch pattern, and the projected row headers are colored dark turquoise with a diagonal hatch pattern.

Variable*	*P*-value	Hazard Ratio (95% CI)	
Analysis without 12-gene risk score			
Gender (Male)	0.22	1.34	(0.84–2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95–2.74)
Tumor Stage			
Stage II	6.25E-05	2.91	(1.72–4.91)
Stage III	1.09E-05	4.16	(2.20–7.85)
Analysis with 12-gene risk score			
Gender (Male)	0.17	1.40	(0.87–2.26)
Age at diagnosis (>60)	0.29	1.34	(0.78–2.31)
Tumor Stage			
Stage II	3.47E-04	2.61	(1.54–4.43)
Stage III	7.40E-06	4.31	(2.28–8.16)
12-gene risk score	**3.10E-05**	**3.94**	**(2.07–7.52)**

(a) Box annotations inferred by the baseline model (TATR v1.1 with bug fixes). The header, which in reality consists only of the top row, is incorrectly inferred to span the top two rows. Of the three spanning cell annotations (green with back-diagonal hatch), only the top-right one is correct.

Variable*	*P*-value	Hazard Ratio (95% CI)	
Analysis without 12-gene risk score			
Gender (Male)	0.22	1.34	(0.84–2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95–2.74)
Tumor Stage			
Stage II	6.25E-05	2.91	(1.72–4.91)
Stage III	1.09E-05	4.16	(2.20–7.85)
Analysis with 12-gene risk score			
Gender (Male)	0.17	1.40	(0.87–2.26)
Age at diagnosis (>60)	0.29	1.34	(0.78–2.31)
Tumor Stage			
Stage II	3.47E-04	2.61	(1.54–4.43)
Stage III	7.40E-06	4.31	(2.28–8.16)
12-gene risk score	**3.10E-05**	**3.94**	**(2.07–7.52)**

(b) The cell matrix produced by the baseline model, corresponding to (a).

Figure 8. The same table image as in Figure 7, processed with the baseline model (TATR v1.1 with bug fixes). The column header is shown in dark orange and with a dotted hatch. The projected row headers are shown in dark turquoise with a diagonal hatch.

8 Future Directions

As noted in Section 7.2, the TATR cell-matrix extraction procedure detects invalid spanning cell annotations in approximately 1.4% of the PubTables-1M complex tables. Because other complex tables may contain additional types of errors, curating the entire subset of complex tables would be a valuable step toward improving both the training and benchmarking of TSR models.

The splitting of the common expression "Hazard Ratio" into separate simple cells (Figure 7(b)) suggests that integrating even basic text-understanding techniques could further enhance performance.

The box relaxation technique presented here could also be applied to more general object detection datasets, particularly as a method for handling noisy or partial annotations.

In cases with missing box annotations, the current requirement to count objects by category can be relaxed, allowing all unannotated objects to be counted

Variable*	*P*-value	Hazard Ratio (95% CI)*
Analysis without 12-gene risk score		
Gender (Male)	0.22	1.34	(0.84–2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95–2.74)
Tumor Stage		
Stage II	6.25E-05	2.91	(1.72–4.91)
Stage III	1.09E-05	4.16	(2.20–7.85)
Analysis with 12-gene risk score		
Gender (Male)	0.17	1.40	(0.87–2.26)
Age at diagnosis (>60)	0.29	1.34	(0.78–2.31)
Tumor Stage		
Stage II	3.47E-04	2.61	(1.54–4.43)
Stage III	7.40E-06	4.31	(2.28–8.16)
12-gene risk score	**3.10E-05**	**3.94**	**(2.07–7.52)**

(a) Box annotations inferred by the constrained box relaxation model. The column header, projected row headers, rows, and columns are correctly identified, with only minor boundary shifts which do not influence the cell matrix extraction. However, the horizontal spanning cell in the top-right corner is not detected.

Variable*	*P*-value	Hazard	Ratio (95% CI)*
Analysis without 12-gene risk score			
Gender (Male)	0.22	1.34	(0.84–2.16)
Age at diagnosis (>60)	0.08	1.61	(0.95–2.74)
Tumor Stage			
Stage II	6.25E-05	2.91	(1.72–4.91)
Stage III	1.09E-05	4.16	(2.20–7.85)
Analysis with 12-gene risk score			
Gender (Male)	0.17	1.40	(0.87–2.26)
Age at diagnosis (>60)	0.29	1.34	(0.78–2.31)
Tumor Stage			
Stage II	3.47E-04	2.61	(1.54–4.43)
Stage III	7.40E-06	4.31	(2.28–8.16)
12-gene risk score	**3.10E-05**	**3.94**	**(2.07–7.52)**

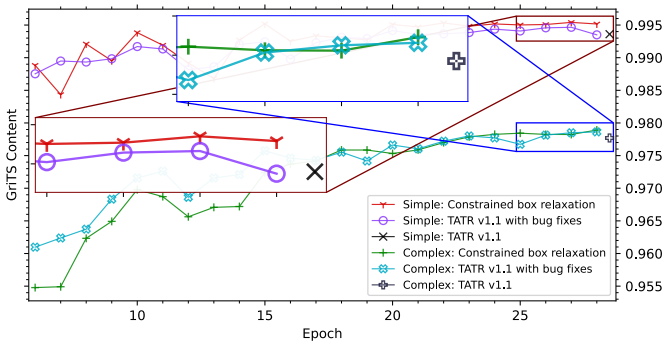
(b) The cell matrix corresponding to (a). The only error is the missing top-right spanning cell, causing the text "Hazard" to appear in its own simple cell.

Figure 9. The same table image as in Figures 7 and 8, processed with the model trained with constrained box relaxation. The column header is shown in dark orange with a dotted hatch. The projected row headers are shown in dark turquoise with a diagonal hatch.

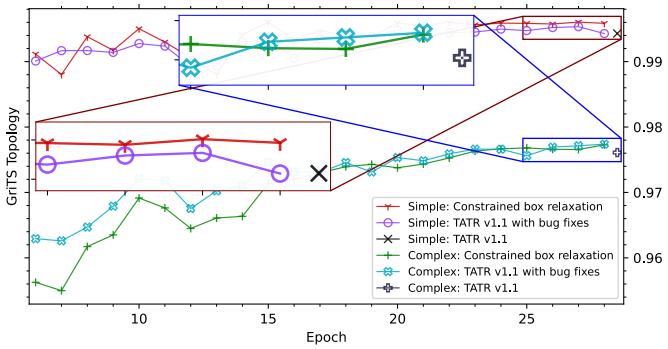
together, regardless of their category.

Extending the method beyond axis-aligned boxes to structured convex shapes, using a technique similar to that of MGIoU [24], would significantly broaden its applicability.

Finally, Appendix C shows that the L_1 distance in center–size format implicitly prioritizes box size, and it proposes the center–half-size format to mitigate this imbalance for object detection with the L_1 distance. A promising direction for future work is to assess



(a) GriTS Content



(b) GriTS Topology

Figure 10. GriTS metrics for simple and complex tables, comparing the constrained box relaxation TSR model to TATR v1.1 with bug fixes. The first five epochs are omitted. The previously published TATR v1.1 model is shown for the best epoch reported in Figure 7 of [12].

the impact of this modification, as well as other use–case-specific linear transformations informed by Appendix B, on DETR and related models.

9 Conclusion

This work demonstrates that box relaxation lowers the barrier to image annotation for training object detection models by requiring bounding boxes for only a subset of objects in each image and by allowing approximate bounding boxes, all without introducing contradictions into the loss function. Furthermore, discretized post-training steps can be leveraged during training as relaxed ground-truth boxes.

Taken together, these results establish box relaxation as an effective tool for reducing dataset annotation requirements. Moreover, in computer vision pipelines with non-injective or discretizing post-inference steps, it can enhance end-to-end performance, as demonstrated for TSR.

Data Availability Statement

The source code used to generate the results reported

here is available under an open-source license at github.com/aioaneid/table-transformer. No new training data were collected for this study.

Funding

This work was supported without any funding.

Acknowledgments

Figures 3, 4, 5, 6, and 7 were adapted from [9, 25–28], respectively, with a better resolution than in PubTables-1M, and annotations added to highlight object bounding boxes or the cell matrix. All sources are licensed under the Creative Commons Attribution (CC BY) license. The author acknowledges BlueCare AG, 8400 Winterthur for computing resources.

Conflicts of Interest

The author declares no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Cai, D., Zhang, Z., & Zhang, Z. (2023). Corner-point and foreground-area IoU loss: Better localization of small objects in bounding box regression. *Sensors, 23*(10), 4961. [CrossRef]

[2] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In *European conference on computer vision* (pp. 213-229). Cham: Springer International Publishing. [CrossRef]

[3] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence, 32*(9), 1627-1645. [CrossRef]

[4] Gou, L., Wu, S., Yang, J., Yu, H., & Li, X. (2022). Gaussian guided IoU: A better metric for balanced learning on object detection. *IET Computer Vision, 16*(6), 556-566. [CrossRef]

[5] He, J., Erfani, S., Ma, X., Bailey, J., Chi, Y., & Hua, X. S. (2021). α-IoU: A family of power intersection over union losses for bounding box regression. *Advances in neural information processing systems, 34*, 20230-20242. [CrossRef]

[6] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Cham: Springer International Publishing. [CrossRef]

[7] Liu, C., Wang, K., Lu, H., Cao, Z., & Zhang, Z. (2022, October). Robust object detection with inaccurate bounding boxes. In *European Conference on Computer Vision* (pp. 53-69). Cham: Springer Nature Switzerland. [CrossRef]

[8] Ma, S., & Xu, Y. (2023). Mpdiou: a loss for efficient and accurate bounding box regression. *arXiv preprint arXiv:2307.07662*.

[9] Pivato, M., De Franceschi, G., Tosatto, L., Frare, E., Kumar, D., Aioanei, D., ... & Bubacco, L. (2012). Covalent α-synuclein dimers: chemico-physical and aggregation properties. *PloS one, 7*(12), e50027. [CrossRef]

[10] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019, June). Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 658-666). IEEE. [CrossRef]

[11] Smock, B., Pesala, R., & Abraham, R. (2022, June). PubTables-1M: Towards comprehensive table extraction from unstructured documents. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4624-4632). IEEE. [CrossRef]

[12] Smock, B., Pesala, R., & Abraham, R. (2023, August). Aligning benchmark datasets for table structure recognition. In *International Conference on Document Analysis and Recognition* (pp. 371-386). Cham: Springer Nature Switzerland. [CrossRef]

[13] Smock, B., Pesala, R., & Abraham, R. (2023, August). GriTS: Grid table similarity metric for table structure recognition. In *International Conference on Document Analysis and Recognition* (pp. 535-549). Cham: Springer Nature Switzerland. [CrossRef]

[14] Suzuki, S. (1985). Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing, 30*(1), 32-46. [CrossRef]

[15] Tang, Y., Wang, J., Wang, X., Gao, B., Dellandréa, E., Gaizauskas, R., & Chen, L. (2017). Visual and semantic knowledge transfer for large scale semi-supervised object detection. *IEEE transactions on pattern analysis and machine intelligence, 40*(12), 3045-3058. [CrossRef]

[16] Tychsen-Smith, L., & Petersson, L. (2018, June). Improving Object Localization with Fitness NMS and Bounded IoU Loss. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6877-6885). IEEE. [CrossRef]

[17] Uma, A., Fornaciari, T., Hovy, D., Paun, S., Plank, B., & Poesio, M. (2020, October). A Case for Soft Loss Functions. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 8, pp. 173-177). [CrossRef]

[18] Wu, Z., Bodla, N., Singh, B., Najibi, M., Chellappa, R., & Davis, L. S. (2020). Soft sampling for robust object detection. In *30th British Machine Vision Conference,*

BMVC 2019.

[19] Zhang, X., Wan, F., Liu, C., Ji, R., & Ye, Q. (2019). Freeanchor: Learning to match anchors for visual object detection. *Advances in neural information processing systems, 32*.

[20] Zhang, X., Yang, Y., & Feng, J. (2019). Learning to localize objects with noisy labeled instances. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI'19)*, pp. 1–10. AAAI Press. [CrossRef]

[21] Xu, J., Ma, Y., He, S., & Zhu, J. (2019). 3D-GIoU: 3D generalized intersection over union for object detection in point cloud. *Sensors, 19*(19), 4093. [CrossRef]

[22] Zhang, Z., & Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems, 31*.

[23] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020, April). Distance-IoU loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 07, pp. 12993-13000). [CrossRef]

[24] Le, D. T., Pham, T., Cai, J., & Rezatofighi, H. (2025). Marginalized Generalized IoU (MGIoU): A Unified Objective Function for Optimizing Any Convex Parametric Shapes. *arXiv preprint arXiv:2504.16443*.

[25] Xiao, C., Li, Y., Baytas, I. M., Zhou, J., & Wang, F. (2018). An MCEM framework for drug safety signal detection and combination from heterogeneous real world evidence. *Scientific reports, 8*(1), 1806. [CrossRef]

[26] García-Minguillán López, O., Jiménez Valbuena, A., & Maestu Unturbe, C. (2019). Significant cellular viability dependence on time exposition at ELF-EMF and RF-EMF in vitro studies. *International journal of environmental research and public health, 16*(12), 2085. [CrossRef]

[27] Aadal, L., Pallesen, H., Arntzen, C., & Moe, S. (2018). Municipal Cross-Disciplinary Rehabilitation following Stroke in Denmark and Norway: A Qualitative Study. *Rehabilitation Research and Practice, 2018*(1), 1972190. [CrossRef]

[28] Wan, Y. W., Sabbagh, E., Raese, R., Qian, Y., Luo, D., Denvir, J., ... & Guo, N. L. (2010). Hybrid models identified a 12-gene signature for lung cancer prognosis and chemoresponse prediction. *PLoS One, 5*(8), e12222. [CrossRef]

A Appendix

B Relaxed L_1 Distance of Transformed Axis-Aligned Boxes (R_1^t)

An algorithm is derived for computing the relaxed L_1 distance between axis-aligned boxes whose coordinates are transformed via function pairs $f_i^{p,q}$: $\mathbb{R}^2 \mapsto \mathbb{R}$, from which the center–size format will follow as a special case.

A *transformed* axis-aligned box in $\mathbb{R}^d$ is represented as a vector $\tilde{b} = (\tilde{l}, \tilde{u}) \in \mathbb{R}^{2d}$, where $\tilde{l}_i = f_i^s(l_i, u_i)$ and $\tilde{u}_i = f_i^t(l_i, u_i)$ for $i = 1..d$.

For a *partially specified* axis-aligned box, if $\tilde{l}_i$ or $\tilde{u}_i$ is missing, both values are treated as missing. That is, for any dimension i, either both $\tilde{l}_i$ and $\tilde{u}_i$ are specified, or neither is.

Definition 3 (R_p^t: Relaxed L_p distance of transformed axis-aligned boxes)**.** *Given*

- *an arbitrary predicted box* $B = (l^b, u^b)$ *and its transformed version* $\tilde{B} = (\tilde{l}^b, \tilde{u}^b)$*, where* $\tilde{l}_i^b = f_i^s(l_i^b, u_i^b)$ *and* $\tilde{u}_i^b = f_i^t(l_i^b, u_i^b)$ *for* $i = 1..d$*, referred to as the* transformed predicted box,
- *a partially specified outer border* $O = (l^o, u^o)$ *and its transformed version* $\tilde{O} = (\tilde{l}^o, \tilde{u}^o)$*, where* $\tilde{l}_i^o = f_i^s(l_i^o, u_i^o)$ *and* $\tilde{u}_i^o = f_i^t(l_i^o, u_i^o)$ *for* $i = 1..d$*, referred to as the* transformed outer border*, and*
- *a partially specified hole border* $H = (l^h, u^h)$ *and its transformed version* $\tilde{H} = (\tilde{l}^h, \tilde{u}^h)$*, where* $\tilde{l}_i^h = f_i^s(l_i^h, u_i^h)$ *and* $\tilde{u}_i^h = f_i^t(l_i^h, u_i^h)$ *for* $i = 1..d$*, referred to as the* transformed hole border,

R_p^t *is defined as*

$$R_p^t(\tilde{B}, \tilde{H}, \tilde{O}) = \min_{H \subseteq I \subseteq O} L_p(\tilde{B}, \tilde{I}) \tag{3}$$

where $0 \leq p \leq \infty$ *and* L_p *denotes the distance induced by the* p*-norm in* $\mathbb{R}^{2d}$*.*

It is shown below that for $p = 1$, the simultaneous distance minimization of f^s and f^t can, for a large class of functions, be simplified to a sequential minimization performed in a specific order. To that end, several auxiliary lemmas are established first.

Lemma 1. *Let* $\mathbb{D} \subseteq \mathbb{E}$ *be two arbitrary sets, and let* $f^{s,t} : \mathbb{E} \mapsto \mathbb{R}$ *be functions such that* $f^t(\mathbb{D})$ *(i.e. the image of* $\mathbb{D}$ *under* f^t*) is a closed interval and,* $\forall X, Y \in \mathbb{D}$*,* $\exists X' \in \mathbb{D}$ s.t. $f^t(X') = f^t(X)$ *and* $|f^s(X') - f^s(Y)| \leq |f^t(X) - f^t(Y)|$*.*

Then for any $B \in \mathbb{E}$ s.t. $\min_{I \in \mathbb{D}} L_1(\tilde{B}, \tilde{I})$ *exists, it holds that*

$$\min_{I \in \mathbb{D}} L_1(\tilde{B}, \tilde{I}) = q + \min_{J \in \mathbb{D} s.t. |f^t(B) - f^t(J)| = q} |f^s(B) - f^s(J)|, \tag{4}$$

where $q = \min_{I \in \mathbb{D}} |f^t(B) - f^t(I)|$*.*

Lemma 1 essentially states that if f^s can change in absolute value no faster than f^t, then the nearest L_1 neighbour of B under the transformation (f^s, f^t) can be found by first minimizing the absolute difference to $f^t(\mathbb{D})$ and then, constrained by that, minimizing the absolute difference to f^s.

Proof of Lemma 1. For an arbitrary $B \in \mathbb{E}$, let $J = J(B) \in \mathbb{D}$ be the point that minimizes the absolute difference to $f^t(B)$, i.e. $|f^t(B) - f^t(J)| = \min_{I\in\mathbb{D}} |f^t(B) - f^t(I)|$. For any $K \in \mathbb{D}$, according to the lemma hypothesis $\exists J' = J'(J(B), K) \in \mathbb{D}$ *s.t.* $f^t(J') = f^t(J)$ and $|f^s(J') - f^s(K)| \leq |f^t(J) - f^t(K)|$. It will be shown next that $L_1(\tilde{B}, \tilde{J'}) \leq L_1(\tilde{B}, \tilde{K})$.

By definition of J, the closest point to $f^t(B)$ in the closed interval $f^t(\mathbb{D})$ is $f^t(J)$. Since $K \in \mathbb{D}$ hence $f^t(K) \in f^t(\mathbb{D})$, it follows that either $f^t(J) = f^t(B)$, or $f^t(K) \leq f^t(J) \leq f^t(B)$, or $f^t(B) \leq f^t(J) \leq f^t(K)$. In all cases, $|f^t(B) - f^t(J)| = |f^t(B) - f^t(K)| - |f^t(K) - f^t(J)|$.

Hence $L_1(\tilde{B}, \tilde{J'}) = |f^s(B) - f^s(J')| + |f^t(B) - f^t(J')| = |f^s(B) - f^s(J')| + |f^t(B) - f^t(J)| = |f^s(B) - f^s(K) + f^s(K) - f^s(J')| + |f^t(B) - f^t(K)| - |f^t(K) - f^t(J)|$. It follows that $L_1(\tilde{B}, \tilde{J'}) \leq |f^s(B) - f^s(K)| + |f^s(K) - f^s(J')| + |f^t(B) - f^t(K)| - |f^t(K) - f^t(J)| \leq |f^s(B) - f^s(K)| + |f^t(B) - f^t(K)| = L_1(\tilde{B}, \tilde{K})$.

Since $K \in \mathbb{D}$ was arbitrary, substituting $I = I(B) = \arg\min_{I\in\mathbb{D}} |f^t(B) - f^t(I)|$ for K shows that the overall minimum is achieved not only at I, but also at $J' = J'(J(B), I(B))$. The latter satisfies $f^t(J') = f^t(J)$, concluding the proof. □

Lemma 2. *Let $\alpha^{s,t}, \beta^{s,t} \in \mathbb{R}$ satisfy $|\alpha^s| \leq |\alpha^t|$ and $|\beta^s| \leq |\beta^t|$. Then, for any $a, b \in \mathbb{R}$, $\exists v = v(a, b), w = w(a, b) \in \mathbb{R}$* s.t. $\min(0, a) \leq v \leq \max(0, a)$, $\min(0, b) \leq w \leq \max(0, b)$, *$\alpha^t v + \beta^t w = \alpha^t a + \beta^t b$, and $|\alpha^s v + \beta^s w| \leq |\alpha^t a + \beta^t b|$.*

Proof of Lemma 2. If $\alpha^t a \beta^t b \geq 0$, then $|\alpha^s a + \beta^s b| \leq |\alpha^s a| + |\beta^s b| = |\alpha^s| \cdot |a| + |\beta^s| \cdot |b| \leq |\alpha^t| \cdot |a| + |\beta^t| \cdot |b| = |\alpha^t a + \beta^t b|$, so the lemma holds with $v = a$ and $w = b$.

Otherwise let $\alpha^t a \beta^t b < 0$. If $|\beta^t b| \geq |\alpha^t a|$, choose $v = 0$ and $w = \frac{\alpha^t}{\beta^t} a + b$, which satisfies $\min(0, b) \leq w \leq \max(0, b)$. Finally let $|\beta^t b| < |\alpha^t a|$ and choose $v = a + \frac{\beta^t}{\alpha^t} b$, which satisfies $\min(0, a) \leq v \leq \max(0, a)$ and $w = 0$, thus concluding the constructive proof. □

Finally, it is shown that, for a large class of linear function pairs, clamping to the image of one function first yields a correct nearest compatible neighbor.

Lemma 3. *Let $f^{s,t} : \mathbb{R} \to \mathbb{R}$ be two real-valued linear functions defined by $f^s(x, y) = \alpha^s x + \beta^s y$, $f^t(x, y) = \alpha^t x + \beta^t y$, with $\alpha^{s,t}, \beta^{s,t} \in \mathbb{R}$. If $|\alpha^s| \leq |\alpha^t|$ and $|\beta^s| \leq |\beta^t|$, then $f^{s,t}$ satisfy the conditions of Lemma 1 for any cross product of closed intervals $\mathbb{D} = [l^o, l^h] \times [u^h, u^o] \subseteq \mathbb{R}^2 = \mathbb{E}$.*

Proof of Lemma 3. Let $X = (l^x, u^x)$ and $Y = (l^y, u^y)$ be points in $\mathbb{D}$ as in Lemma 1. Apply Lemma 2 with $a = l^x - l^y$, $b = u^x - u^y$, to obtain $v = v(X, Y)$, $w = w(X, Y)$. Define $X' = (l^{x'}, u^{x'})$, $l^{x'} = l^y + v$, $u^{x'} = u^y + w$.

It will be shown next that the four conditions of Lemma 1 are satisfied by X, Y, and X':

- $f^t(\mathbb{D})$ is a closed interval;
- $X' \in \mathbb{D}$;
- $f^t(X') = f^t(X)$;
- $|f^s(X') - f^s(Y)| \leq |f^t(X) - f^t(Y)|$.

The first condition follows from the linearity of f^t: $f^t(\mathbb{D}) = [\min(f^t(\mathbb{S})), \max(f^t(\mathbb{S}))]$ where $\mathbb{S} = \{f^t(l^o, u^h), f^t(l^o, u^o), f^t(l^h, u^h), f^t(l^h, u^o)\}$.

The second condition is satisfied because $l^o \leq \min(l^x, l^y) = l^y + \min(0, l^x - l^y) \leq l^{x'} \leq l^y + \max(0, l^x - l^y) = \max(l^x, l^y) \leq l^h$ and similarly $u^h \leq \min(u^x, u^y) = u^y + \min(0, u^x - u^y) \leq u^{x'} \leq u^y + \max(0, u^x - u^y) = \max(u^x, u^y) \leq u^o$.

The third condition can be proved as follows: $f^t(X') = \alpha^t l^{x'} + \beta^t u^{x'} = \alpha^t(l^y + v) + \beta^t(u^y + w) = (\alpha^t l^y + \beta^t u^y) + (\alpha^t v + \beta^t w) = (\alpha^t l^y + \beta^t u^y) + (\alpha^t a + \beta^t b) = (\alpha^t l^y + \beta^t u^y) + \alpha^t(l^x - l^y) + \beta^t(u^x - u^y) = \alpha^t l^x + \beta^t u^x = f^t(X)$.

Finally the fourth condition follows from $|f^s(X') - f^s(Y)| = |\alpha^s l^{x'} + \beta^s u^{x'} - (\alpha^s l^y + \beta^s u^y)| = |\alpha^s(l^{x'} - l^y) + \beta^s(u^{x'} - u^y)| = |\alpha^s v + \beta^s w| \leq |\alpha^t a + \beta^t b| = |\alpha^t(l^x - l^y) + \beta^t(u^x - u^y)| = |(\alpha^t l^x + \beta^t u^x) - (\alpha^t l^y + \beta^t u^y)| = |f^t(X) - f^t(Y)|$, which concludes the constructive proof. □

The complete algorithm for the general case of Lemma 3 is given in Algorithm 2. The algorithm for the center–size format, as used in DETR, is obtained as the special case with $\alpha^s = \beta^s = ½$ and $\alpha^t = -1$, $\beta^t = 1$.

Proof of Algorithm 2. The dimensions are independent of each other, so it is sufficient to prove correctness in the 1-dimensional case. To simplify notation, the dimension index i is dropped.

If the outer border is missing, it is set equal to the full range of the respective dimension, *i.e.*, no localization information is provided from the outer

Algorithm 2: R_1^t: L_1 with transformed coordinates

Input: $\tilde{B} = (\tilde{l}^b, \tilde{u}^b), \tilde{O} = (\tilde{l}^o, \tilde{u}^o), \tilde{H} = (\tilde{l}^h, \tilde{u}^h)$, $\alpha^s_{1..d}, \beta^s_{1..d}, \alpha^t_{1..d}, \beta^t_{1..d}$;
$|\alpha^s_i| \leq |\beta^s_i|, 0 < |\alpha^t_i| \leq |\beta^t_i|, \alpha^s_i\beta^t_i \neq \beta^s_i\alpha^t_i, i = 1..d$
for $i = 1..d$ **do**
  $\delta_i \leftarrow \alpha^s_i\beta^t_i - \beta^s_i\alpha^t_i$; $\gamma_i \leftarrow \alpha^s_i\alpha^t_i - \beta^s_i\beta^t_i$
  if $\delta_i\alpha^t_i > 0$ **then**
    if $\delta_i\beta^t_i > 0$ **then**
      $\tilde{l}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^o_i + \beta^t_i\tilde{l}^o_i) + \beta^t_i(\alpha^s_i\tilde{u}^h_i - \alpha^t_i\tilde{l}^h_i)$;
      $\tilde{u}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^h_i + \beta^t_i\tilde{l}^h_i) + \beta^t_i(\alpha^s_i\tilde{u}^o_i - \alpha^t_i\tilde{l}^o_i)$
    else
      $\tilde{l}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^o_i + \beta^t_i\tilde{l}^o_i) + \beta^t_i(\alpha^s_i\tilde{u}^o_i - \alpha^t_i\tilde{l}^o_i)$;
      $\tilde{u}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^h_i + \beta^t_i\tilde{l}^h_i) + \beta^t_i(\alpha^s_i\tilde{u}^h_i - \alpha^t_i\tilde{l}^h_i)$
    end if
  else
    if $\delta_i\beta^t_i > 0$ **then**
      $\tilde{l}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^h_i + \beta^t_i\tilde{l}^h_i) + \beta^t_i(\alpha^s_i\tilde{u}^h_i - \alpha^t_i\tilde{l}^h_i)$;
      $\tilde{u}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^o_i + \beta^t_i\tilde{l}^o_i) + \beta^t_i(\alpha^s_i\tilde{u}^o_i - \alpha^t_i\tilde{l}^o_i)$
    else
      $\tilde{l}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^h_i + \beta^t_i\tilde{l}^h_i) + \beta^t_i(\alpha^s_i\tilde{u}^o_i - \alpha^t_i\tilde{l}^o_i)$;
      $\tilde{u}^t_i \leftarrow \alpha^t_i(-\beta^s_i\tilde{u}^o_i + \beta^t_i\tilde{l}^o_i) + \beta^t_i(\alpha^s_i\tilde{u}^h_i - \alpha^t_i\tilde{l}^h_i)$
    end if
  end if
  $\tilde{u}_i \leftarrow \min(\max(\delta_i\tilde{u}^b_i, \tilde{l}^t_i), \tilde{u}^t_i)$ {Clamp $\delta_i\tilde{u}^b_i$}
  if $\alpha^t_i > 0$ **then**
    if $\beta^t_i > 0$ **then**
      $\tilde{l}^s_i \leftarrow \max(\tilde{u}_i\beta^s_i\alpha^t_i + \delta_i\alpha^t_i(\beta^t_i\tilde{l}^o_i - \beta^s_i\tilde{u}^o_i), \tilde{u}_i\alpha^s_i\beta^t_i - \delta_i\beta^t_i(-\alpha^t_i\tilde{l}^o_i + \alpha^s_i\tilde{u}^o_i))$
      $\tilde{u}^s_i \leftarrow \min(\tilde{u}_i\beta^s_i\alpha^t_i + \delta_i\alpha^t_i(\beta^t_i\tilde{l}^h_i - \beta^s_i\tilde{u}^h_i), \tilde{u}_i\alpha^s_i\beta^t_i - \delta_i\beta^t_i(-\alpha^t_i\tilde{l}^h_i + \alpha^s_i\tilde{u}^h_i))$
    else
      $\tilde{l}^s_i \leftarrow \max(\tilde{u}_i\beta^s_i\alpha^t_i + \delta_i\alpha^t_i(\beta^t_i\tilde{l}^o_i - \beta^s_i\tilde{u}^o_i), \tilde{u}_i\alpha^s_i\beta^t_i - \delta_i\beta^t_i(-\alpha^t_i\tilde{l}^h_i + \alpha^s_i\tilde{u}^h_i))$
      $\tilde{u}^s_i \leftarrow \min(\tilde{u}_i\beta^s_i\alpha^t_i + \delta_i\alpha^t_i(\beta^t_i\tilde{l}^h_i - \beta^s_i\tilde{u}^h_i), \tilde{u}_i\alpha^s_i\beta^t_i - \delta_i\beta^t_i(-\alpha^t_i\tilde{l}^o_i + \alpha^s_i\tilde{u}^o_i))$
    end if
  else
    if $\beta^t_i > 0$ **then**
      $\tilde{l}^s_i \leftarrow \max(\tilde{u}_i\beta^s_i\alpha^t_i + \delta\alpha^t_i(\beta^t_i\tilde{l}^h_i - \beta^s_i\tilde{u}^h_i), \tilde{u}_i\alpha^s_i\beta^t_i - \beta^t_i\delta(-\alpha^t_i\tilde{l}^o_i + \alpha^s_i\tilde{u}^o_i))$
      $\tilde{u}^s_i \leftarrow \min(\tilde{u}_i\beta^s_i\alpha^t_i + \delta\alpha^t_i(\beta^t_i\tilde{l}^o_i - \beta^s_i\tilde{u}^o_i), \tilde{u}_i\alpha^s_i\beta^t_i - \beta^t_i\delta(-\alpha^t_i\tilde{l}^h_i + \alpha^s_i\tilde{u}^h_i))$
    else
      $\tilde{l}^s_i \leftarrow \max(\tilde{u}_i\beta^s_i\alpha^t_i + \delta\alpha^t_i(\beta^t_i\tilde{l}^h_i - \beta^s_i\tilde{u}^h_i), \tilde{u}_i\alpha^s_i\beta^t_i - \beta^t_i\delta(-\alpha^t_i\tilde{l}^h_i + \alpha^s_i\tilde{u}^h_i))$
      $\tilde{u}^s_i \leftarrow \min(\tilde{u}_i\beta^s_i\alpha^t_i + \delta\alpha^t_i(\beta^t_i\tilde{l}^o_i - \beta^s_i\tilde{u}^o_i), \tilde{u}_i\alpha^s_i\beta^t_i - \beta^t_i\delta(-\alpha^t_i\tilde{l}^o_i + \alpha^s_i\tilde{u}^o_i))$
    end if
  end if
  $\tilde{l}_i \leftarrow \min(\max(\alpha^t_i\beta^t_i\delta\tilde{l}^b_i, \tilde{l}^s_i), \tilde{u}^s_i)$ {Clamp $\delta_i\alpha^t_i\beta^t_i\tilde{l}^b_i$}
end for
$\tilde{X} \leftarrow (\tilde{l}/(\delta\alpha^t\beta^t), \tilde{u}/\delta)$ {Nearest neighbor}
return $||\tilde{B} - \tilde{X}||_1$ {Optimal distance}

border. Similarly, a missing hole border interval is removed by setting $l^h = u^o$ and $u^h = l^o$. In this way, both hole border constraints $l^o \leq l \leq l^h$ and $u^h \leq u \leq u^o$ become equivalent to $l^o \leq l \leq u^o = l^h$ and $u^h = l^o \leq u \leq u^o$, respectively.

By Lemma 3, it is valid to first clamp $\tilde{u}^b$ to its valid range. Multiplying by the non-zero constant δ, $\delta\tilde{u}^b$ can then be clamped to the interval $[\tilde{l}^t, \tilde{u}^t]$, where $\tilde{l}^t = \min_{l^o \leq l \leq l^h, u^h \leq u \leq u^o} \delta(\alpha^t l + \beta^t u)$ and $\tilde{u}^t = \max_{l^o \leq l \leq l^h, u^h \leq u \leq u^o} \delta(\alpha^t l + \beta^t u)$.

In this formulation, l^o, u^o, l^h, u^h are not directly provided, but $\tilde{l}^{o,h} = \alpha^s l^{o,h} + \beta^s u^{o,h}$ and $\tilde{u}^{o,h} = \alpha^t l^{o,h} + \beta^t u^{o,h}$ are given instead. Solving this system gives $\delta l^{o,h} = -\beta^s\tilde{u}^{o,h} + \beta^t\tilde{l}^{o,h}$ and $\delta u^{o,h} = \alpha^s\tilde{u}^{o,h} - \alpha^t\tilde{l}^{o,h}$.

Observe that $\delta(\alpha^t l + \beta^t u) = \alpha^t(\delta l) + \beta^t(\delta u)$. When $\delta\alpha^t > 0$ and $\delta\beta^t > 0$, the minimum becomes $\tilde{l}^t = \alpha^t(\delta l^o) + \beta^t(\delta u^h)$ and the maximum $\tilde{u}^t = \alpha^t(\delta l^h) + \beta^t(\delta u^o)$, exactly as computed by the algorithm. The other cases are analogous.

Once $\delta\tilde{u}^b$ is clamped to obtain $\tilde{u}$, that can be enforced as the constraint $\delta(\alpha^t l + \beta^t u) = \tilde{u}$, which is equivalent to $\delta\beta^t u = \tilde{u} - \delta\alpha^t l$. Then $\delta\alpha^t\beta^t(\alpha^s l + \beta^s u) = \delta\alpha^s\alpha^t\beta^t l + \beta^s\alpha^t(\tilde{u} - \delta\alpha^t l) = \beta^s\alpha^t\tilde{u} + \delta\alpha^t(\delta l)$. If $\alpha^t > 0$, the condition $l^o \leq l \leq l^h$ becomes $\beta^s\alpha^t\tilde{u} + \delta\alpha^t(-\beta^s\tilde{u}^o + \beta^t\tilde{l}^o) = \beta^s\alpha^t\tilde{u} + \delta\alpha^t(\delta l^o) \leq \delta\alpha^t\beta^t(\alpha^s l + \beta^s u) \leq \beta^s\alpha^t\tilde{u} + \delta\alpha^t(\delta l^h) = \beta^s\alpha^t\tilde{u} + \delta\alpha^t(-\beta^s\tilde{u}^h + \beta^t\tilde{l}^h)$. The case $\alpha^t < 0$ is treated similarly. By writing $\delta\alpha^t l = \tilde{u} - \delta\beta^t u$ the condition $u^h \leq u \leq u^o$ yields an analogous interval constraint. The algorithm clamps $\delta\alpha^t\beta^t\tilde{l}^b$ to the intersection of these two intervals, which concludes the correctness proof. □

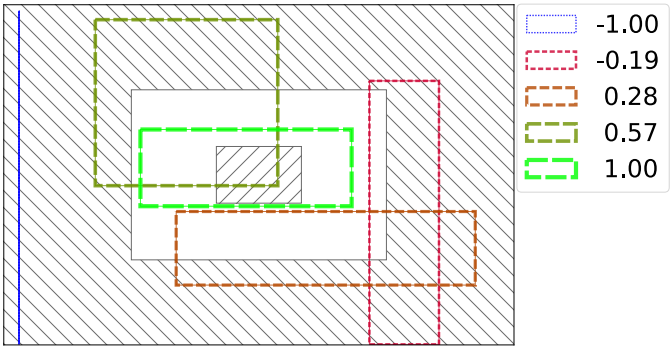
C Center–size Representation

Figure A1 shows 2D examples of $R_1^t(\cdot, \cdot, \cdot, ½, ½, -1, 1)$.

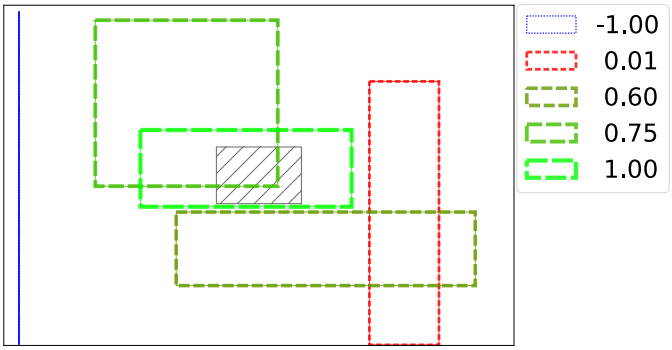
A comparison of the nearest compatible neighbours in lower- and upper-bound corner coordinates (R_1) and the center–size format ($R_1^t(\cdot, \cdot, \cdot, ½, ½, -1, 1)$) is shown in Figure A2.

Interestingly, Figure A2 can be interpreted outside the context of box relaxation by considering the predicted boxes and their nearest neighbors to be target boxes and predicted boxes, respectively. More precisely, for each (target) box B, among the two nearest-neighbor (predicted) boxes, the L_1 distance in center–size format is smaller for the predicted box that better preserves the size.

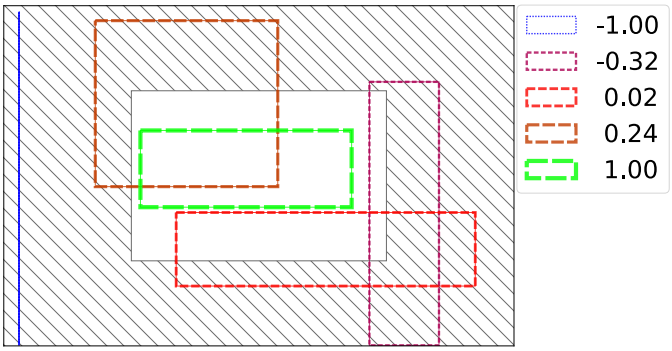
Section Appendix B can aid in designing alternative



(a) Expression $1 - 2R_1^t(\tilde{B}, \tilde{H}, \tilde{O}, ½, ½, -1, 1) / \max_B R_1^t(\tilde{B}, \tilde{H}, \tilde{O}, ½, ½, -1, 1)$ for rectangles B relative to hole and outer borders $H \subseteq O$. R_1^t reaches its minimum value 0 if and only if $H \subseteq B \subseteq O$.



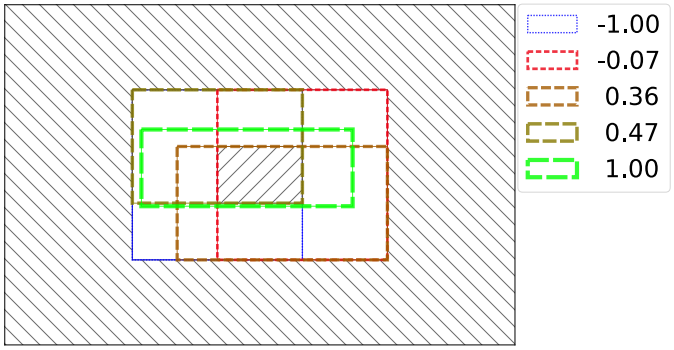
(b) Expression $1 - 2R_1^t(\tilde{B}, \tilde{H}, \_, ½, ½, -1, 1) / \max_B R_1^t(\tilde{B}, \tilde{H}, \_, ½, ½, -1, 1)$ for rectangles B relative to hole border H. R_1^t reaches its minimum value 0 if and only if $H \subseteq B$.



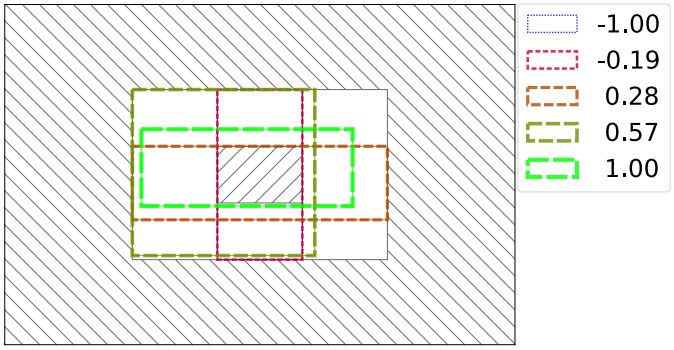
(c) Expression $1 - 2R_1^t(\tilde{B}, \_, \tilde{O}, ½, ½, -1, 1) / \max_B R_1^t(\tilde{B}, \_, \tilde{O}, ½, ½, -1, 1)$ for rectangles B relative to outer border O. R_1^t reaches its minimum value 0 if and only if $B \subseteq O$.

Figure A1. Expression $1 - 2R_1^t(\tilde{B}, \cdot, \cdot, ½, ½, -1, 1) / \max_B R_1^t(\tilde{B}, \cdot, \cdot, ½, ½, -1, 1)$ for axis-aligned boxes in 2D. A tilde on an axis-aligned box variable denotes its representation in the center–size format. Penalty areas are hatched. The color palette and the boxes B, hole border H, and outer border O are as in Figures 1 and 2.

linear transformations with different preferences. For instance, to achieve a better balance between box



(a) Nearest compatible neighbor according to $R_1(B, H, O)$ from Figure A1(b), computed using Algorithm 1. Lower- and upper-bound corners are preserved as much as possible.



(b) Nearest compatible neighbor according to $R_1^t(\tilde{B}, \tilde{H}, \tilde{O}, ½, ½, -1, 1)$ from Figure A1(c), computed using Algorithm 2. Box sizes are preserved as much as possible, whereas centers play a secondary role.

Figure A2. Comparison of $R_1(B, H, O)$ and $R_1^t(\tilde{B}, \tilde{H}, \tilde{O}, ½, ½, -1, 1)$ for rectangles B relative to hole border H and outer border O. The nearest compatible neighbour depends on the representation of the axis-aligned boxes. A tilde on a box variable indicates the center–size representation. Penalty areas are hatched. The color palette and the boxes B, hole border H, and outer border O are as in Figures 1, 2 and A1.

center and size, the size should be considered halved, resulting in the center–half-size format. This ensures that $|\alpha^s| = |\alpha^t| = ½$ and $|\beta^s| = |\beta^t| = ½$, so that Lemma 3 remains equally valid if f^s and f^t in Definition 3 are swapped.

D Constrained Box Relaxation for Table Structure Recognition (TSR)

Given the image box, text spans, and original ground-truth boxes in Equation (5), where the two dimensions are denoted by $i = 0..1$, these represent a known feasible interior point for the non-linear optimization problem in Equation (6), with $\mathcal{L}(\text{box})$ denoting the perimeter of an axis-aligned box.

Algorithm 3: $TsrRelax$: Relax target table object boxes under GriTS equivalence

$T^h \leftarrow T$; $C^h_j \leftarrow C_j, j = 1..n_c$; $R^h_j \leftarrow R_j, j = 1..n_r$;
$K^h_j \leftarrow K_j, j = 1..n_k$; $P^h_j \leftarrow P_j, j = 1..n_p$; $S^h_j \leftarrow S_j, j = 1..n_s$;
{Hole borders equal to input boxes}
$T^o \leftarrow T$; $C^o_j \leftarrow C_j, j = 1..n_c$; $R^o_j \leftarrow R_j, j = 1..n_r$;
$K^o_j \leftarrow K_j, j = 1..n_k$; $P^o_j \leftarrow P_j, j = 1..n_p$; $S^o_j \leftarrow S_j, j = 1..n_s$;
{Outer borders equal to input boxes}
while not converged **do**
  for $b \in [1024, 512, 256, 128, 64, 32, 16, 8, 4, 2, 1]$ **do**
    Generate a random permutation of size $8 \times (1 + n_c + n_r + n_k + n_p + n_s)$ of all relaxed corner coordinates;
    for each relaxed corner coordinate in the random permutation order: **do**
      $\alpha \leftarrow$ uniform random number in $[0, b)$;
      if this is a lower bound of a hole border, or an upper bound of an outer border **then**
        Add α to the current relaxed corner coordinate;
      else
        Subtract α from the current relaxed corner coordinate.
      end if
      If any constraint is violated, revert the last update $\pm\alpha$.
    end for
  end for
end while
return T^h, T^o

$$
\begin{aligned}
\text{Image: } I &= (l^z_0, l^z_1, u^z_0, u^z_1); \\
l^z_i &\le u^z_i, i = 0..1 \\
\text{Text spans: } A_j &= (l^a_{j,0}, l^a_{j,1}, u^a_{j,0}, u^a_{j,1}) \subseteq I; \\
l^a_{j,i} &\le u^a_{j,i}, j = 1..n_a, i = 0..1 \\
\text{Table (T.): } T &= (l^t_0, l^t_1, u^t_0, u^t_1) \subseteq I; \\
l^z_i &\le l^t_i \le u^t_i \le u^z_i, i = 0..1 \\
\text{Optional T. col. header: } K_j &= (l^k_{j,0}, l^k_{j,1}, u^k_{j,0}, u^k_{j,1}) \subseteq T; \\
l^k_{j,i} &\le u^k_{j,i}, j = 1..n_k, \\
n_k &\in \{0, 1\}, i = 0..1, \\
\text{T. proj. row headers: } P_j &= (l^p_{j,0}, l^p_{j,1}, u^p_{j,0}, u^p_{j,1}) \subseteq T; \\
l^p_{j,i} &\le u^p_{j,i}, j = 1..n_p, i = 0..1 \\
\text{T. spanning cells: } S_j &= (l^s_{j,0}, l^s_{j,1}, u^s_{j,0}, u^s_{j,1}) \subseteq T; \\
l^s_{j,i} &\le u^s_{j,i}, j = 1..n_s, i = 0..1 \\
\text{T. columns: } C_j &= (l^c_{j,0}, l^c_{j,1}, u^c_{j,0}, u^c_{j,1}) \subseteq T; \\
l^c_{j,i} &\le u^c_{j,i}, j = 1..n_c, i = 0..1 \\
&\text{T. columns intersecting T. proj. row headers} \\
&\text{or T. spanning cells are sorted left-to-right.} \\
\text{T. rows: } R_j &= (l^r_{j,0}, l^r_{j,1}, u^r_{j,0}, u^r_{j,1}) \subseteq T; \\
l^r_{j,i} &\le u^r_{j,i}, j = 1..n_r, i = 0..1 \\
&\text{T. rows intersecting T. proj. row headers} \\
&\text{or T. spanning cells are sorted top-to-bottom.}
\end{aligned}
\tag{5}
$$

This optimization problem is applied independently to each training example, making it straightforward to parallelize. In this work, the randomized approach presented in Algorithm 3 was used to relax the object boxes for TSR.

$$
\begin{aligned}
\underset{\substack{T^h, T^o; \\ C^h_j, C^o_j, j=1..n_c; \\ R^h_j, R^o_j, j=1..n_r; \\ K^h_j, K^o_j, j=1..n_k; \\ P^h_j, P^o_j, j=1..n_p; \\ S^h_j, S^o_j, j=1..n_s}}{\text{maximize}} \quad & \mathcal{L}(T^o) - \mathcal{L}(T^h) \\
& + \sum_{j=1}^{n_c} [\mathcal{L}(C^o_j) - \mathcal{L}(C^h_j)] \\
& + \sum_{j=1}^{n_r} [\mathcal{L}(R^o_j) - \mathcal{L}(R^h_j)] \\
& + \sum_{j=1}^{n_k} [\mathcal{L}(K^o_j) - \mathcal{L}(K^h_j)] \\
& + \sum_{j=1}^{n_p} [\mathcal{L}(P^o_j) - \mathcal{L}(P^h_j)] \\
& + \sum_{j=1}^{n_s} [\mathcal{L}(S^o_j) - \mathcal{L}(S^h_j)] \qquad (6a)
\end{aligned}
$$

subject to

$$
\begin{aligned}
& \text{center}(T) \in T^h \subseteq T \subseteq T^o \subseteq I \\
\wedge\, & \text{center}(C_j) \in C^h_j \subseteq C_j \subseteq C^o_j \subseteq I, \quad j = 1..n_c \\
\wedge\, & \text{center}(R_j) \in R^h_j \subseteq R_j \subseteq R^o_j \subseteq I, \quad j = 1..n_r \\
\wedge\, & \text{center}(K_j) \in K^h_j \subseteq K_j \subseteq K^o_j \subseteq I, \quad j = 1..n_k \\
\wedge\, & \text{center}(P_j) \in P^h_j \subseteq P_j \subseteq P^o_j \subseteq I, \quad j = 1..n_p \\
\wedge\, & \text{center}(S_j) \in S^h_j \subseteq S_j \subseteq S^o_j \subseteq I, \quad j = 1..n_s; \qquad (6b)
\end{aligned}
$$

$$
2|A_j \cap T| < |A_j| \implies 2|A_j \cap T^o| < |A_j|, \quad j = 1..n_a; \qquad (6c)
$$

$$
2|A_j \cap T| \ge |A_j| \implies 2|A_j \cap T^h| \ge |A_j|, \quad j = 1..n_a; \qquad (6d)
$$

Other constraints are omitted for brevity. (6e)

Dr. Daniel Aioanei received his Master's degree in Bioinformatics from the University of Bologna, Italy, in 2009, followed by a PhD in Molecular, Cellular, and Functional Biology in 2013. After a decade-long career as a Software Engineer at Google, he worked in the Swiss healthcare software industry and, most recently, as a Robotics Software Engineer specializing in tethered drones. (Email: aioaneid@gmail.com)