Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression

SUPPLEMENTARY MATERIALS

Hamid Rezatofighi[1,2] Nathan Tsoi[1] JunYoung Gwak[1] Amir Sadeghian[1,3]
Ian Reid[2] Silvio Savarese[1]

[1]Computer Science Department, Stanford University, United states
[2]School of Computer Science, The University of Adelaide, Australia
[3]Aibee Inc, USA
hamidrt@stanford.edu

This supplementary material accompanies our main submission. In an effort to further explore select topics covered in the main text and provide the necessary tools for future experimentation we provide the following. First, we extend $GIoU$ definition to a more general case for both convex and non-convex shapes. Then, we show an analytical solution for $\mathcal{L}_{GIoU}$ as a loss for n-orthotopes. Next, we provide proofs for the $GIoU$ properties described in the main text. Then, we show the derivation of the gradient of $\mathcal{L}_{GIoU}$. Finally, we provide more qualitative object detection results.

1. GIoU extention

In this section, we extend $GIoU$ definition to the arbitrary shapes (convex or non-convex). If both A and B are not convex shapes, the current definition for $\mathcal{L}_{GIoU}$ cannot fulfill the second property of a metric, *i.e.* identity of indiscernibles. Because if $A = B$, C as the smallest enclosing convex shape for A and B will not overlay A and B. Therefore $\mathcal{L}_{GIoU}(A, B) \neq 0$ (Check Sec. 3.3.2 for an intuition). To address this limitation, we extend the definition of $GIoU$, making it applicable to both convex and non-convex shapes.

To generalize $GIoU$, we re-define C to be attained by applying a function, *e.g.* geometric operations, on the one of the shapes, *i.e.* $C = \mathcal{F}_{A,B}(B), \quad \mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$, such that $\mathcal{F}_{A,B}$ ensure that A and B are enclosed by the smallest possible C. The definition for $\mathcal{F}_{A,B}$ can be rather arbitrary; however it should fulfill the following properties:

- The output space of $\mathcal{F}$ should be identical to its argument space, *i.e.* if $C = \mathcal{F}_{A,B}(B)$, then $C, B \subseteq \mathbb{S} \in \mathbb{R}^n$.
- The area (or volume) of C, should be monotonically decreasing when B tends to A by shape similarity and spatial proximity.
- When two objects A and B overlay perfectly, *i.e.* $A = B$, the smallest object C for A and B is exactly A or B, and therefore $|C| = |A| = |B| = |A \cap B| = |A \cup B|$.

Algorithm 1: Extension to Generalized Intersection over Union

```
input  : Given two arbitrary shapes as a reference $A$ and prediction $B$: $A, B \subseteq \mathbb{S} \in \mathbb{R}^n$
output: $GIoU$
For $A$ and $B$, find the smallest enclosing object $C$, where $C \subseteq \mathbb{S} \in \mathbb{R}^n$, which is attained by applying a function, e.g. geometric
  operations, on $B$, i.e. $C = \mathcal{F}_{A,B}(B)$, such that $\mathcal{F}_{A,B}$ ensure that $A$ and $B$ are enclosed by the smallest possible $C$
$IoU = \frac{|A \cap B|}{|A \cup B|}$
$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|}$
```

One of the possible examples for the function $\mathcal{F}_{A,B}$ would be an affine transformation on the predicted shape B. Therefore in this case, C is tightest scaled, translated and rotated version of B, enclosing A and B. This definition is also consistent with what we proposed in the case of axis aligned rectangles, where C can be seen the tightest scaled and translated version of the predicted box, enclosing both bounding boxes.
Compared to the definition proposed in the main text, this extension has two advantages: a) It addresses the limitation of $GIoU$ distance to be a metric when one of the shapes is non-convex, b) C encodes the shape information of the one of the objects which reflects not only their proximity, but also their shape similarity.

2. $\mathcal{L}_{GIoU}$ as a loss for n-orthotopes

In this section, we provide the analytical solution for $\mathcal{L}_{GIoU}$ as a loss for n-orthotopes. Note that any axis-aligned hyperrectangle can be uniquely defined by any two corner coordinates. The subtraction of one coordinate from the other gives the size along a given axis. We therefore use this representation for the following generalization to n-orthotopes.

```
Algorithm 2: $\mathcal{L}_{GIoU}$ as a loss for axis-aligned n-orthotopes
input : The coordinates for two vertices of two axis-aligned n-orthotopes, i.e. the ground truth $\mathcal{O}^g$ and predicted $\mathcal{O}^p$ n-orthotopes:
        $\mathcal{O}^g = (x^g_{1,1}, x^g_{2,1}, \cdots, x^g_{n,1}, x^g_{1,2}, x^g_{2,2}, \cdots, x^g_{n,2}) \subseteq \mathbb{R}^{2n}$,
        $\mathcal{O}^p = (x^p_{1,1}, x^p_{2,1}, \cdots, x^p_{n,1}, x^p_{1,2}, x^p_{2,2}, \cdots, x^p_{n,2}) \subseteq \mathbb{R}^{2n}$.
output: $\mathcal{L}_{GIoU}$.
For the predicted n-orthotope $\mathcal{O}^p$, ensuring $x^p_{1,2} > x^p_{1,1}$, $x^p_{2,2} > x^p_{2,1}$, $\cdots$, and $x^p_{n,2} > x^p_{n,1}$:
$\hat{x}^p_{1,1} = \min(x^p_{1,2}, x^p_{1,1}), \quad \hat{x}^p_{2,1} = \min(x^p_{2,2}, x^p_{2,1}), \quad \cdots, \quad \hat{x}^p_{n,1} = \min(x^p_{n,2}, x^p_{n,1})$,
$\hat{x}^p_{1,2} = \max(x^p_{1,2}, x^p_{1,1}), \quad \hat{x}^p_{2,2} = \max(x^p_{2,2}, x^p_{2,1}), \quad \cdots, \quad \hat{x}^p_{n,2} = \max(x^p_{n,2}, x^p_{n,1})$.
Calculating the volume of the ground truth n-orthotope $\mathcal{O}^g$: $V^g = (x^g_{1,2} - x^g_{1,1}) \times (x^g_{2,2} - x^g_{2,1}) \times \cdots \times (x^g_{n,2} - x^g_{n,1})$.
Calculating the volume of the predicted n-orthotope $\mathcal{O}^p$: $V^p = (\hat{x}^p_{1,2} - \hat{x}^p_{1,1}) \times (\hat{x}^p_{2,2} - \hat{x}^p_{2,1}) \times \cdots \times (\hat{x}^p_{n,2} - \hat{x}^p_{n,1})$.
Calculating intersection $\mathcal{I}$ between $\mathcal{O}^p$ and $\mathcal{O}^g$:
$x^{\mathcal{I}}_{1,1} = \max(\hat{x}^p_{1,1}, x^g_{1,1}), \quad x^{\mathcal{I}}_{2,1} = \max(\hat{x}^p_{2,1}, x^g_{2,1}), \quad \cdots, \quad x^{\mathcal{I}}_{n,1} = \max(\hat{x}^p_{n,1}, x^g_{n,1})$,
$x^{\mathcal{I}}_{1,2} = \min(\hat{x}^p_{1,1}, x^g_{1,1}), \quad x^{\mathcal{I}}_{2,2} = \min(\hat{x}^p_{2,1}, x^g_{2,1}), \quad \cdots, \quad x^{\mathcal{I}}_{n,2} = \min(\hat{x}^p_{n,1}, x^g_{n,1})$,
$\mathcal{I} = \begin{cases} (x^{\mathcal{I}}_{1,2} - x^{\mathcal{I}}_{1,1}) \times (x^{\mathcal{I}}_{2,2} - x^{\mathcal{I}}_{2,1}) \times \cdots \times (x^{\mathcal{I}}_{n,2} - x^{\mathcal{I}}_{n,1}) & \text{if} \quad x^{\mathcal{I}}_{1,2} > x^{\mathcal{I}}_{1,1}, x^{\mathcal{I}}_{2,2} > x^{\mathcal{I}}_{2,1}, \cdots, x^{\mathcal{I}}_{n,2} > x^{\mathcal{I}}_{n,1} \\ 0 & \text{otherwise.} \end{cases}$
Finding the coordinate of smallest enclosing n-orthotope $\mathcal{O}^c$:
$x^c_{1,1} = \min(\hat{x}^p_{1,1}, x^g_{1,1}), \quad x^c_{2,1} = \min(\hat{x}^p_{2,1}, x^g_{2,1}), \quad \cdots, \quad x^c_{n,1} = \min(\hat{x}^p_{n,1}, x^g_{n,1})$,
$x^c_{1,2} = \max(\hat{x}^p_{1,1}, x^g_{1,1}), \quad x^c_{2,2} = \max(\hat{x}^p_{2,1}, x^g_{2,1}), \quad \cdots, \quad x^c_{n,2} = \max(\hat{x}^p_{n,1}, x^g_{n,1})$,
Calculating the volume of $\mathcal{O}^c$: $V^c = (x^c_{1,2} - x^c_{1,1}) \times (x^c_{2,2} - x^c_{2,1}) \times \cdots \times (x^c_{n,2} - x^c_{n,1})$.
$IoU = \frac{\mathcal{I}}{\mathcal{U}}$, where the union $\mathcal{U}$ is attained by $\mathcal{U} = V^p + V^g - \mathcal{I}$.
$GIoU = IoU - \frac{V^c - \mathcal{U}}{V^c}$.
    $\mathcal{L}_{GIoU} = 1 - GIoU$.
```

3. GIoU properties

As briefed in the main text, $GIoU$ has some appealing properties. In this section, we provide the proof for each of these properties.

3.1. Scale invariance

Proposition 1: IoU and $GIoU$ are invariant to the scale of the problem.

Proof 1: In order to prove this proposition, we use the following theorem.

Theorem: For any arbitrary area/volume $\forall A \subseteq \mathbb{S} \in \mathbb{R}^n$, if the space $\mathbb{S}$ is scaled by a scaling factor γ, *i.e.* $|\tilde{\mathbb{S}}| = \gamma|\mathbb{S}|$, any area/volume in this space will be scaled by this scaling factor, *i.e.* $\forall|\tilde{A}| = \gamma|A|$, where $\tilde{A} \subseteq \tilde{\mathbb{S}} \in \mathbb{R}^n$.

Since the intersection $\mathcal{I} = A \cap B \subseteq \mathbb{S}$, union $\mathcal{U} = A \cup B \subseteq \mathbb{S}$, the smallest enclosing object for A and B, $C \subseteq \mathbb{S}$

and $\mathcal{C}_e = C\backslash A \cup B \subseteq \mathbb{S}$ are also arbitrary areas/volumes in the space $\mathbb{S}$, scaling the space, *i.e.* $|\tilde{\mathbb{S}}| = \gamma|\mathbb{S}|$, will scale their volume as well, *i.e.* $|\tilde{\mathcal{I}}| = \gamma|\mathcal{I}|$, $|\tilde{\mathcal{U}}| = \gamma|\mathcal{U}|$, $|\tilde{C}| = \gamma|C|$ and $|\tilde{\mathcal{C}_e}| = \gamma|\mathcal{C}_e|$, where $\tilde{\mathcal{I}}$, $\tilde{\mathcal{U}}$, $\tilde{C}$, $\tilde{\mathcal{C}}_e \subseteq \tilde{\mathbb{S}}$.

Therefore,

$$IoU(\tilde{A},\tilde{B}) = \frac{|\tilde{\mathcal{I}}|}{|\tilde{\mathcal{U}}|} = \frac{\gamma|\mathcal{I}|}{\gamma|\mathcal{U}|} = \frac{|\mathcal{I}|}{|\mathcal{U}|} = IoU(A,B)$$

and similarly,

$$GIoU(\tilde{A},\tilde{B}) = \frac{|\tilde{\mathcal{I}}|}{|\tilde{\mathcal{U}}|} - \frac{|\tilde{\mathcal{C}_e}|}{|\tilde{C}|} = \frac{\gamma|\mathcal{I}|}{\gamma|\mathcal{U}|} - \frac{\gamma|\mathcal{C}_e|}{\gamma|C|} = \frac{|\mathcal{I}|}{|\mathcal{U}|} - \frac{|\mathcal{C}_e|}{|C|} = GIoU(A,B)$$

3.2. GIoU, a lower bound for IoU

Proposition 2: $GIoU$ is always a lower bound for IoU, *i.e.* $\forall A, B \subseteq \mathbb{S}$, $GIoU(A,B) \leq IoU(A,B)$, and this lower bound becomes tighter when A and B have a stronger shape similarity and proximity, *i.e.* $\lim_{A\to B} GIoU(A,B) = IoU(A,B)$.

Proof 2: To prove this proposition, we use the following trivial properties for the smallest enclosing object C for A and B:

- $A \cap B \subseteq A \subseteq A \cup B \subseteq C \subseteq \mathbb{S}$
- $A \cap B \subseteq B \subseteq A \cup B \subseteq C \subseteq \mathbb{S}$

As a result, $|A \cup B| \leq |C| \leq |\mathbb{S}|$, or $0 \leq |C| - |A \cup B| \leq |\mathbb{S}| - |A \cup B|$. Therefore, $0 \leq \frac{|C|-|A\cup B|}{|C|} \leq \frac{|\mathbb{S}|-|A\cup B|}{|C|}$ or $-\frac{|\mathbb{S}|-|A\cup B|}{|C|} \leq -\frac{|C|-|A\cup B|}{|C|} \leq 0$. Consequently, we can add the value $IoU = \frac{|A\cap B|}{|A\cup B|}$ to each term of inequality, *i.e.*:

$$\frac{|A\cap B|}{|A\cup B|} - \frac{|\mathbb{S}|-|A\cup B|}{|C|} \leq \underbrace{\frac{|A\cap B|}{|A\cup B|} - \frac{|C|-|A\cup B|}{|C|}}_{GIoU(A,B)} \leq \underbrace{\frac{|A\cap B|}{|A\cup B|}}_{IoU(A,B)}$$

When A and B have a stronger shape similarity and proximity, the difference between $|A \cup B|$ and $|A \cap B|$ is small. The difference between $|A \cup B|$ and the volume of the smallest enclosing object $|C|$ also needs to be less significant. Otherwise A and B should be far apart from each other to ensure $|C| \gg |A \cup B|$ which contradicts the first argument. Therefore:

$$\lim_{A\to B} GIoU(A,B) = \frac{|A\cap B|}{|A\cup B|} - \frac{\overbrace{|C|-|A\cup B|}^{\approx 0}}{|C|} = \frac{|A\cap B|}{|A\cup B|} = IoU(A,B)$$

Proposition 3: when two objects A and B overlay perfectly, *i.e.* if $|A \cup B| = |A \cap B|$, then $GIoU = 1$. Also $GIoU$ value asymptotically converges to -1 when the ratio between occupying regions of two shapes, $A \cup B$, and the volume (area) of the enclosing shape C tends to zero, *i.e.* $\lim_{\frac{|A\cup B|}{|C|}\to 0} GIoU(A,B) = -1$.

Proof 3: When two objects A and B overlay perfectly, then $|A| = |B| = |A \cap B| = |A \cup B|$. Therefore, the smallest object C for A and B is exactly A or B, *i.e.* $|C| = |A| = |B| = |A \cap B| = |A \cup B|$. Then, we have:

$$GIoU(A,B) = \frac{|A\cap B|}{|A\cup B|} - \frac{|C|-|A\cup B|}{|C|} = \frac{|A\cap B|}{|A|+|B|-|A\cap B|} - \frac{|C|-|A|-|B|+|A\cap B|}{|C|}$$
$$= \frac{|A|}{|A|+|A|-|A|} - \frac{|A|-|A|-|A|+|A|}{|A|} = 1$$

In the other extreme case, when the ratio between occupying regions of two shapes, $A \cup B$, and the volume (area) of the enclosing shape C tends to zero, *i.e.* $|C| \gg |A \cup B|$ or $|C| \gg |A|+|B|-|A \cap B|$. This case is possible, when $|C| \gg |A|+|B|$ and $|A \cap B| \approx 0$. Otherwise if $|A \cap B| \gg 0$, A and B should have a stronger shape similarity and proximity. Therefore according to **Proof 2**, the difference between $|A \cup B|$ and $|C|$ is less significant, which contradicts $|C| \gg |A \cup B|$. Thus, $|C| \gg |A| + |B| - |A \cap B|$ is possible, when $|C| \gg |A| + |B|$ and $|A \cap B| \approx 0$. Consequently, we have:

$$\lim_{\frac{|A\cup B|}{|C|}\to 0} GIoU(A,B) = \frac{\overbrace{|A\cap B|}^{\approx 0}}{|A\cup B|} - \frac{\overbrace{|C|-|A|-|B|}^{\approx |C|}+\overbrace{|A\cap B|}^{\approx 0}}{|C|} = -1$$

3.3. $\mathcal{L}_{GIoU}$ is a metric

The proof to show IoU as a distance, *i.e.* $\mathcal{L}_{IoU} = 1 - IoU$, is a metric has been provided in different articles [1, 3, 4, 5]. In this section, we provide a proof to show $GIoU$ as a distance, *i.e.* $\mathcal{L}_{GIoU} = 1 - GIoU$, holds all properties of a metric such as non-negativity, identity of indiscernibles, symmetry and triangle inequality.

3.3.1 Non-negativity

Proposition 4: For any two shapes A and B, $\mathcal{L}_{GIoU}$ is non-negative, *i.e.* $\forall A, B \subseteq \mathbb{S}$, $\mathcal{L}_{GIoU}(A, B) \geq 0$.

Proof 4: To prove this, we use the aforementioned trivial properties in **Proof 2** for the smallest enclosing object C for A and B, *i.e.* $A \cap B \subseteq A \cup B \subseteq C$. Thus $0 \leq |A \cap B| \leq |A \cup B|$ and $0 \leq |A \cup B| \leq |C|$. Since $|A \cup B| \geq 0$ and $|C| \geq 0$, we can divide each inequality with these positive values, *i.e.* $0 \leq \frac{|A \cap B|}{|A \cup B|} \leq 1$ and $0 \leq \frac{|A \cup B|}{|C|} \leq 1$. Therefore, by summing up these two inequalities, we have:

$$0 \leq \frac{|A \cap B|}{|A \cup B|} + \frac{|A \cup B|}{|C|} \leq 2.$$

We then multiply -1 to each side of inequality:

$$-2 \leq -\frac{|A \cap B|}{|A \cup B|} - \frac{|A \cup B|}{|C|} \leq 0.$$

Next, we add 2 to each side of inequality:

$$0 \leq 2 - \frac{|A \cap B|}{|A \cup B|} - \frac{|A \cup B|}{|C|} \leq 2 \Rightarrow 0 \leq 1 - \frac{|A \cap B|}{|A \cup B|} + 1 - \frac{|A \cup B|}{|C|} \leq 2.$$

$$\Rightarrow 0 \leq \underbrace{1 - \frac{|A \cap B|}{|A \cup B|} + \frac{|C| - |A \cup B|}{|C|}}_{\mathcal{L}_{GIoU}} \leq 2$$

Consequently, $\mathcal{L}_{GIoU} \geq 0$.

3.3.2 Identity of indiscernibles

Proposition 5: $\mathcal{L}_{GIoU}(A, B) = 0 \Leftrightarrow B = A$.

Proof 5: As we discussed in **Proof 3**, when two objects A and B overlay perfectly, *i.e.* $A = B$, the smallest object C for A and B is exactly A or B, and therefore $|C| = |A| = |B| = |A \cap B| = |A \cup B|$.

$$\text{if } A = B \Rightarrow \mathcal{L}_{GIoU}(A, B) = 1 - \frac{\overbrace{|A \cap B|}^{A \text{or} B}}{\underbrace{|A \cup B|}_{A \text{or} B}} + \overbrace{\frac{|C| - |A \cup B|}{|C|}}^{=0} = 1 - 1 + 0 = 0$$

To show if $\mathcal{L}_{GIoU}(A, B) = 0 \Rightarrow A = B$, we can prove this by contradiction, *i.e.*

$$\text{if } \mathcal{L}_{GIoU}(A, B) = 0 \Rightarrow 1 - \frac{|A \cap B|}{|A \cup B|} + \frac{|C| - |A \cup B|}{|C|} = 0 \Rightarrow \frac{|A \cap B|}{|A \cup B|} - \frac{|C| - |A \cup B|}{|C|} = 1$$

if $A \neq B$, then $A \cup B \neq A \cap B$. Thus, $0 \leq \frac{|A \cap B|}{|A \cup B|} < 1$. Moreover, as discussed in **Proof 2**, $-\frac{|C| - |A \cup B|}{|C|} \leq 0$. Therefore,

$$\text{if } A \neq B \Rightarrow \frac{|A \cap B|}{|A \cup B|} - \frac{|C| - |A \cup B|}{|C|} < 1$$

which contradicts $\frac{|A \cap B|}{|A \cup B|} - \frac{|C| - |A \cup B|}{|C|} = 1$. Consequently, $\mathcal{L}_{GIoU}(A, B) = 0 \Rightarrow A = B$.

3.3.3 Symmetry

Proposition 6: For any two convex shapes A and B, $\mathcal{L}_{GIoU}(A,B) = \mathcal{L}_{GIoU}(B,A)$.

Proof 6: Based on commutative laws of set algebra, we have $A \cup B = B \cup A$ and $A \cap B$=$B \cap A$. Moreover, the smallest enclosing convex hull (shape) between any A and B is a symmetric function and does not depend on the order of A and B, *i.e.* $C_{A,B} = C_{B,A}$. Accordingly, we have:

$$\mathcal{L}_{GIoU}(B,A) = 1 - \frac{|B \cap A|}{|B \cup A|} + \frac{|C_{B,A}| - |B \cup A|}{|C_{B,A}|} = 1 - \frac{|A \cap B|}{|A \cup B|} + \frac{|C_{A,B}| - |A \cup B|}{|C_{A,B}|} = \mathcal{L}_{GIoU}(A,B)$$

3.3.4 Triangle inequality

Proposition 7: For any three shapes A_i, A_j and A_k, triangle inequality holds true, *i.e.*

$$\mathcal{L}_{GIoU}(A_i,A_k) \leq \mathcal{L}_{GIoU}(A_i,A_j) + \mathcal{L}_{GIoU}(A_j,A_k).$$

[1] **Proof 7a (special cases):**
Case 1: If none of the pairs from A_i, A_j and A_k overlap, *i.e.* $|A_i \cap A_k| = |A_j \cap A_k| = |A_i \cap A_j| = 0$:
Since $0 \leq \frac{|A_i \cup A_k|}{|C_{i,k}|}, \frac{|A_i \cup A_j|}{|C_{i,j}|}, \frac{|A_j \cup A_k|}{|C_{j,k}|} \leq 1$, the following inequality also holds true:

$$2 + \frac{|A_i \cup A_k|}{|C_{i,k}|} \geq \frac{|A_i \cup A_j|}{|C_{i,j}|} + \frac{|A_j \cup A_k|}{|C_{j,k}|}.$$

By multiplying -1 and adding 4 to each side of the inequality, we have:

$$2 - \frac{|A_i \cup A_k|}{|C_{i,k}|} \leq 2 - \frac{|A_i \cup A_j|}{|C_{i,j}|} + 2 - \frac{|A_j \cup A_k|}{|C_{j,k}|}.$$

Since $|A_i \cap A_k| = |A_j \cap A_k| = |A_i \cap A_j| = 0$, we can subtract some zero values, *e.g.* the terms $\frac{|A_i \cap A_k|}{|A_i \cup A_k|} = \frac{|A_j \cap A_k|}{|A_j \cup A_k|} = \frac{|A_i \cap A_j|}{|A_i \cup A_j|} = 0$, from each side of the inequality:

$$2 - \frac{|A_i \cap A_k|}{|A_i \cup A_k|} - \frac{|A_i \cup A_k|}{|C_{i,k}|} \leq 2 - \frac{|A_i \cap A_j|}{|A_i \cup A_j|} - \frac{|A_i \cup A_j|}{|C_{i,j}|} + 2 - \frac{|A_j \cap A_k|}{|A_j \cup A_k|} - \frac{|A_j \cup A_k|}{|C_{j,k}|}.$$

$$1 - \frac{|A_i \cap A_k|}{|A_i \cup A_k|} + 1 - \frac{|A_i \cup A_k|}{|C_{i,k}|} \leq 1 - \frac{|A_i \cap A_j|}{|A_i \cup A_j|} + 1 - \frac{|A_i \cup A_j|}{|C_{i,j}|} + 1 - \frac{|A_j \cap A_k|}{|A_j \cup A_k|} + 1 - \frac{|A_j \cup A_k|}{|C_{j,k}|}.$$

$$\underbrace{1 - \frac{|A_i \cap A_k|}{|A_i \cup A_k|} + \frac{|C_{i,k}| - |A_i \cup A_k|}{|C_{i,k}|}}_{\mathcal{L}_{GIoU}(A_i,A_k)} \leq \underbrace{1 - \frac{|A_i \cap A_j|}{|A_i \cup A_j|} + \frac{|C_{i,j}| - |A_i \cup A_j|}{|C_{i,j}|}}_{\mathcal{L}_{GIoU}(A_i,A_j)} + \underbrace{1 - \frac{|A_j \cap A_k|}{|A_j \cup A_k|} + \frac{|C_{j,k}| - |A_j \cup A_k|}{|C_{j,k}|}}_{\mathcal{L}_{GIoU}(A_j,A_k)}.$$

Case 2: If one of the pairs overlay perfectly, *e.g.* $|A_i \cup A_k| = |A_i \cap A_k| = |C_{i,k}|$:
In this case,

$$\mathcal{L}_{GIoU}(A_i,A_k) = 2 - \underbrace{\frac{|A_i \cap A_k|}{|A_i \cup A_k|}}_{=1} - \underbrace{\frac{|A_i \cup A_k|}{|C_{i,k}|}}_{=1} = 0,$$

[1] Deriving the exact proof appears to be far from straightforward. Therefore, we show the validity of this proposition by 1) an exact proof for special cases, 2) experimentally, checking 10^6 random samples for any counterexample.

and also $|A_j \cup A_k| = |A_i \cup A_j|$, $|A_j \cap A_k| = |A_i \cap A_j|$, $|C_{j,k}| = |C_{i,j}|$. Therefore, we have:

$$\mathcal{L}_{GIoU}(A_i, A_j) + \mathcal{L}_{GIoU}(A_j, A_k) = 4 - 2 \times \underbrace{\left(\frac{|A_i \cap A_j|}{|A_i \cup A_j|} + \frac{|A_i \cup A_j|}{|C_{i,j}|}\right)}_{\leq 2} \geq \mathcal{L}_{GIoU}(A_i, A_k).$$

Case 3: If only one of the shapes are very far from the other two, *e.g.* $|A_i \cap A_k| = 0$, $|A_i \cap A_j| = 0$, $|A_i \cup A_k| \ll |C_{i,k}|$ and $|A_i \cup A_j| \ll |C_{i,j}|$:
In this case, we have:

$$\mathcal{L}_{GIoU}(A_i, A_k) = 2 - \underbrace{\frac{|A_i \cap A_k|}{|A_i \cup A_k|}}_{=0} - \underbrace{\frac{|A_i \cup A_k|}{|C_{i,k}|}}_{\approx 0} \approx 2,$$

$$\mathcal{L}_{GIoU}(A_i, A_j) = 2 - \underbrace{\frac{|A_i \cap A_j|}{|A_i \cup A_j|}}_{=0} - \underbrace{\frac{|A_i \cup A_j|}{|C_{i,j}|}}_{\approx 0} \approx 2,$$

and

$$0 \leq \mathcal{L}_{GIoU}(A_j, A_k) = 2 - \frac{|A_j \cap A_k|}{|A_j \cup A_k|} - \frac{|A_j \cup A_k|}{|C_{j,k}|} \leq 2,$$

Therefore, the triangle inequality holds true in this case:

$$\mathcal{L}_{GIoU}(A_i, A_k) \leq \mathcal{L}_{GIoU}(A_i, A_j) + \mathcal{L}_{GIoU}(A_j, A_k).$$

Case 4: For general case, we check the correctness of the proposition by evaluating many random samples, detailed next.

Proof 7b (random sampling):
Over 10^6 iterations we sample 3 convex hulls, denoted at H_i, H_j, H_k, each composed of 4 points, where each point is represented by its (x, y) coordinate. For each pair of elements in this randomly sampled set of 3 convex hulls, we compute $\mathcal{L}_{GIoU}$, e.g. $\mathcal{L}_{GIoU}(H_i, H_k)$, $\mathcal{L}_{GIoU}(H_i, H_j)$, and $\mathcal{L}_{GIoU}(H_j, H_k)$. To compute the intersection for each pair of convex hulls, as required to compute $\mathcal{L}_{GIoU}$, we check for intersection points between any pair of the line segments that compose each convex hull. Each point on both convex hulls in the pair is checked to determine if the point is enclosed in the other convex hull. Intersection of two hulls is calculated by forming a new convex hull: the tightest convex hull of the enclosed points and the intersection points as determined above. C is calculated by forming the tightest convex hull of all points. The following condition is then tested: $\mathcal{L}_{GIoU}(H_i, H_k) \leq \mathcal{L}_{GIoU}(H_i, H_j) + \mathcal{L}_{GIoU}(H_j, H_k)$. Throughout all samples the above condition held.

4. *Gradient* of $\mathcal{L}_{GIoU}$

In this section we provide the gradient of $\mathcal{L}_{GIoU}$. Following the derivation of $\mathcal{L}_{GIoU}$ we include the derivations necessary to account for different bounding box representations.

4.1. Notation

We follow the same notation as the main text. Given bounding box B, where B^g denotes ground truth and B^p denotes a prediction, $B = (x_1, y_1, x_2, y_2)$ and (x_1, y_1) is the upper left corner of the bounding box and (x_2, y_2) is the lower right corner of the bounding box, or in vector form:

$$B = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix} \tag{1}$$

4.2. $\mathcal{L}_{GIoU}$ Gradient

Let $\frac{\partial GIoU}{\partial x_1 \partial y_1 \partial x_2 \partial y_2}$ denote the derivative of GIoU with respect to x_1, y_1, x_2, y_2 or in the general case, simply: $\frac{\partial GIoU}{\partial B}$.

Let ∇_{x_1,y_1,x_2,y_2} indicate the gradient function with respect to x_1, y_1, x_2, y_2 or generally, ∇_B.

Intersection $\mathcal{I}$ between prediction bounding box B^p and ground truth bounding box B^g is given by:

$$\begin{aligned} x_1^{\mathcal{I}} &= \max(\hat{x}_1^p, x_1^g), \quad x_2^{\mathcal{I}} = \min(\hat{x}_2^p, x_2^g), \\ y_1^{\mathcal{I}} &= \max(\hat{y}_1^p, y_1^g), \quad y_2^{\mathcal{I}} = \min(\hat{y}_2^p, y_2^g) \end{aligned} \tag{2}$$

$$\begin{aligned} \mathcal{I}_h &= (y_2^{\mathcal{I}} - y_1^{\mathcal{I}}) \\ \mathcal{I}_w &= (x_2^{\mathcal{I}} - x_1^{\mathcal{I}}) \end{aligned} \tag{3}$$

$$\mathcal{I} = \begin{cases} \mathcal{I}_h \times \mathcal{I}_w & \text{if} \quad \mathcal{I}_h > 0, \mathcal{I}_h > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

$\nabla_B \mathcal{I}$ is given by:

$$\frac{\partial \mathcal{I}}{\partial x_1} = \left(\mathcal{I}_w\right)\left(0\right) + \begin{cases} -1 & \hat{x}_1^p > x_1^g \\ 0 & otherwise \end{cases} \left(\mathcal{I}_h\right) \tag{5}$$

$$\frac{\partial \mathcal{I}}{\partial y_1} = \begin{cases} -1 & \hat{y}_1^p > y_1^g \\ 0 & otherwise \end{cases} \left(\mathcal{I}_w\right) + \left(\mathcal{I}_h\right)\left(0\right) \tag{6}$$

$$\frac{\partial \mathcal{I}}{\partial x_2} = \left(\mathcal{I}_w\right)\left(0\right) + \begin{cases} 1 & \hat{x}_2^p < x_2^g \\ 0 & otherwise \end{cases} \left(\mathcal{I}_h\right) \tag{7}$$

$$\frac{\partial \mathcal{I}}{\partial y_2} = \begin{cases} 1 & \hat{y}_2^p < y_2^g \\ 0 & otherwise \end{cases} \left(\mathcal{I}_w\right) + \left(\mathcal{I}_h\right)\left(0\right) \tag{8}$$

Recall the smallest enclosing box B^c is given by:

$$\begin{aligned} x_1^c &= \min(\hat{x}_1^p, x_1^g), \quad x_2^c = \max(\hat{x}_2^p, x_2^g), \\ y_1^c &= \min(\hat{y}_1^p, y_1^g), \quad y_2^c = \max(\hat{y}_2^p, y_2^g) \end{aligned} \tag{9}$$

Area of B^c, called A^c is given by:

$$\begin{aligned} A_h^c &= (y_2^c - y_1^c) \\ A_w^c &= (x_2^c - x_1^c) \end{aligned} \tag{10}$$

$$A^c = A_h^c \times A_w^c \tag{11}$$

$\nabla_B A^c$ is given by:

$$\frac{\partial A^c}{\partial x_1} = \left(A_w^c\right)\left(0\right) + \begin{cases} -1 & \hat{x}_1^p < x_1^g \\ 0 & otherwise \end{cases} \left(A_h^c\right) \tag{12}$$

$$\frac{\partial A^c}{\partial y_1} = \begin{cases} -1 & \hat{y}_1^p < y_1^g \\ 0 & otherwise \end{cases} \left(A_w^c\right) + \left(A_h^c\right)\left(0\right) \tag{13}$$

$$\frac{\partial A^c}{\partial x_2} = \left(A_w^c\right)\left(0\right) + \begin{cases} 1 & \hat{x}_2^p > x_2^g \\ 0 & otherwise \end{cases} \left(A_h^c\right) \tag{14}$$

$$\frac{\partial A^c}{\partial y_2} = \begin{cases} 1 & \hat{y}_2^p > y_2^g \\ 0 & otherwise \end{cases} \left(A_w^c\right) + \left(A_h^c\right)\left(0\right) \tag{15}$$

Recall:

$$GIoU = \frac{\mathcal{I}}{\mathcal{U}} - \frac{A^c - \mathcal{U}}{A^c} \tag{16}$$

Therefore:

$$\frac{\partial GIoU}{dx} = \begin{cases} \frac{\mathcal{U}(\nabla_B \mathcal{I}) - \mathcal{I}(\nabla_B \mathcal{U})}{\mathcal{U}^2} & if \quad \mathcal{I}_h > 0, \mathcal{I}_w > 0 \\ 0 & \text{otherwise.} \end{cases} - \frac{A^c(\nabla_B \mathcal{U}) - \mathcal{U}(\nabla_B A^c)}{{A^c}^2} \tag{17}$$

Where:

$$\nabla_B \mathcal{U} = \nabla_B A^g + \nabla_B A^p - \nabla_B \mathcal{I} \tag{18}$$

Recall that prediction area is given by:

$$A^p = (\hat{y}_2^p - \hat{y}_1^p) \times (\hat{x}_2^p - \hat{x}_1^p) \tag{19}$$

Ground truth area is given by:

$$A^g = (\hat{y}_2^g - \hat{y}_1^g) \times (\hat{x}_2^g - \hat{x}_1^g) \tag{20}$$

Generally, the element-wise derivative for area A (A^g or A^p), for point $(\hat{x}, \hat{y})$ [$(\hat{x}^g$,$\hat{y}^g)$ or $(\hat{x}^p$, $\hat{y}^p)$], is given by:

$$\frac{\partial A}{x_1} = (\hat{y}_2 - \hat{y}_1)(-1) + (0)(\hat{x}_2 - \hat{x}_1) = \hat{y}_1 - \hat{y}_2 \tag{21}$$

$$\frac{\partial A}{y_1} = (\hat{y}_2 - \hat{y}_1)(0) + (-1)(\hat{x}_2 - \hat{x}_1) = \hat{x}_1 - \hat{x}_2 \tag{22}$$

$$\frac{\partial A}{x_2} = (\hat{y}_2 - \hat{y}_1)(1) + (0)(\hat{x}_2 - \hat{x}_1) = \hat{y}_2 - \hat{y}_1 \tag{23}$$

$$\frac{\partial A}{y_2} = (\hat{y}_2 - \hat{y}_1)(0) + (1)(\hat{x}_2 - x_1) = \hat{x}_2 - \hat{x}_1 \tag{24}$$

Therefore:

$$\begin{bmatrix} \frac{\partial GIoU}{dx_1} \\ \frac{\partial GIoU}{dy_1} \\ \frac{\partial GIoU}{dx_2} \\ \frac{\partial GIoU}{dy_2} \end{bmatrix} = \begin{bmatrix} \frac{U(\frac{\partial \mathcal{I}}{\partial x_1}) - \mathcal{I}(\frac{\partial A^p}{\partial x_1} - \frac{\partial \mathcal{I}}{\partial x_1})}{U^2} - \frac{A^c(\frac{\partial A^p}{\partial x_1} - \frac{\partial \mathcal{I}}{\partial x_1}) - \mathcal{U}(\frac{\partial A^c}{\partial x_1})}{A^c_2} \\ \frac{U(\frac{\partial \mathcal{I}}{\partial y_1}) - \mathcal{I}(\frac{\partial A^p}{\partial y_1} - \frac{\partial \mathcal{I}}{\partial y_1})}{U^2} - \frac{A^c(\frac{\partial A^p}{\partial y_1} - \frac{\partial \mathcal{I}}{\partial y_1}) - \mathcal{U}(\frac{\partial A^c}{\partial y_1})}{A^c_2} \\ \frac{U(\frac{\partial \mathcal{I}}{\partial x_2}) - \mathcal{I}(\frac{\partial A^p}{\partial x_2} - \frac{\partial \mathcal{I}}{\partial x_2})}{U^2} - \frac{A^c(\frac{\partial A^p}{\partial x_2} - \frac{\partial \mathcal{I}}{\partial x_2}) - \mathcal{U}(\frac{\partial A^c}{\partial x_2})}{A^c_2} \\ \frac{U(\frac{\partial \mathcal{I}}{\partial y_2}) - \mathcal{I}(\frac{\partial A^p}{\partial y_2} - \frac{\partial \mathcal{I}}{\partial y_2})}{U^2} - \frac{A^c(\frac{\partial A^p}{\partial y_2} - \frac{\partial \mathcal{I}}{\partial y_2}) - \mathcal{U}(\frac{\partial A^c}{\partial y_2})}{A^c_2} \end{bmatrix} \tag{25}$$

4.3. Handling Different Output Representations

In cases where the predicted bounding box B^p, is represented by the top left and bottom right coordinates and it is not guaranteed to adhere to $x_2^p > x_1^p$ and $y_2^p > y_1^p$, we apply:

$$\begin{gathered} \hat{x}_1^p = \min(x_1^p, x_2^p), \quad \hat{x}_2^p = \max(x_1^p, x_2^p), \\ \hat{y}_1^p = \min(y_1^p, y_2^p), \quad \hat{y}_2^p = \max(y_1^p, y_2^p) \end{gathered} \tag{26}$$

In these cases, we use the following partial derivatives:

$$\frac{\partial \hat{x}_1^p}{\partial x_1^p} = \begin{cases} 1 & x_1^p < x_2^p \\ 0 & otherwise \end{cases} \quad \frac{\partial \hat{x}_1^p}{\partial x_2^p} = \begin{cases} 1 & x_1^p > x_2^p \\ 0 & otherwise \end{cases} \tag{27}$$

$$\frac{\partial \hat{x}_2^p}{\partial x_1^p} = \begin{cases} 1 & x_1^p > x_2^p \\ 0 & otherwise \end{cases} \quad \frac{\partial \hat{x}_2^p}{\partial x_2^p} = \begin{cases} 1 & x_1^p < x_2^p \\ 0 & otherwise \end{cases} \tag{28}$$

$$\frac{\partial \hat{y}_1^p}{\partial y_1^p} = \begin{cases} 1 & y_1^p < y_2^p \\ 0 & otherwise \end{cases} \qquad \frac{\partial \hat{y}_1^p}{\partial y_2^p} = \begin{cases} 1 & y_1^p > y_2^p \\ 0 & otherwise \end{cases} \tag{29}$$

$$\frac{\partial \hat{y}_2^p}{\partial y_1^p} = \begin{cases} 1 & y_1^p > y_2^p \\ 0 & otherwise \end{cases} \qquad \frac{\partial \hat{y}_2^p}{\partial y_2^p} = \begin{cases} 1 & y_1^p < y_2^p \\ 0 & otherwise \end{cases} \tag{30}$$

Alternatively, for networks that represent bounding boxes with (x, y, w, h) where (x, y) is the coordinate of the top-left corner and w, h are width and height respectively, the following additional step during back propagation may be necessary – a coordinate space transformation from the network's representation to the $B = (x_1, y_1, x_2, y_2)$ representation. This is given by:

$$T(x, y, w, h) = \begin{bmatrix} x_x - \frac{x_w}{2} \\ x_y - \frac{x_h}{2} \\ x_x + \frac{x_w}{2} \\ x_y + \frac{x_h}{2} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix} \tag{31}$$

The Jacobian of T is given by:

$$J(x, y, w, h) = \begin{bmatrix} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & -\frac{1}{2} \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & \frac{1}{2} \end{bmatrix} \tag{32}$$

Transpose of the Jacobian is given by:

$$J^T(x, y, w, h) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \tag{33}$$

Therefore:

$$\begin{bmatrix} \partial x \\ \partial y \\ \partial w \\ \partial h \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{\partial GIoU}{\partial x_1} \\ \frac{\partial GIoU}{\partial y_1} \\ \frac{\partial GIoU}{\partial x_2} \\ \frac{\partial GIoU}{\partial y_2} \end{bmatrix} \tag{34}$$

5. Further qualitative results

In this section we show further qualitative results. In general we believe that qualitative results are well correlated with our quantitative results showing that $\mathcal{L}_{GIoU}$ is superior to $\mathcal{L}_{IoU}$ which is superior to MSE loss for localization. However, in some cases we see that $\mathcal{L}_{GIoU}$ and $\mathcal{L}_{IoU}$ are not able to perform classification as well as the baseline losses due to sub-optimal normalization between the new localization loss and classification loss.

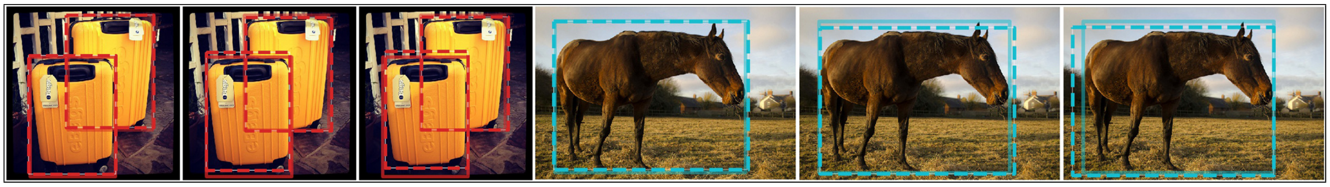

Figure 1. Example results from COCO validation using Mask R-CNN [2] trained using (left to right) $\mathcal{L}_{GIoU}$, $\mathcal{L}_{IoU}$, ℓ_1-smooth losses. Ground truth is shown by a solid line and predictions are represented with dashed lines.

Fig. 2 shows a result from both Mask R-CNN [2] and YOLO v3 [6] where $GIoU$ exhibits the strong localization accuracy, but has a decreased classification score. Also, notice the localization accuracy vs classification score between YOLO v3 and

Mask R-CNN across all losses. As noted below, some dashed outlines are almost transparent since the opacity of these dashed line is set to the network output score for a given prediction bounding box. Solid white lines indicate ground truth.

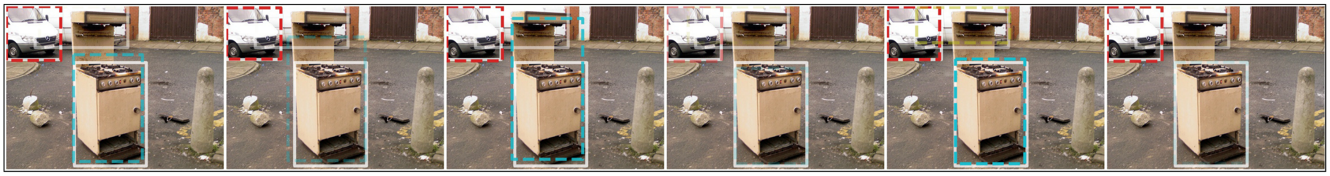

Figure 2. Example results from COCO validation for both YOLO v3 [6] (left) and Mask R-CNN [2] (right) trained using (left to right): $\mathcal{L}_{GIoU}$, $\mathcal{L}_{IoU}$ and MSE for YOLO v3 and $\mathcal{L}_{GIoU}$, $\mathcal{L}_{IoU}$ and ℓ_1-smooth for Mask R-CNN. Ground truth is shown by a solid white line and predictions are represented with a dashed line. The opacity of the prediction corresponds to the confidence of the prediction.

Fig. 3 shows two cases where $GIoU$ exhibits increased localization accuracy, however classification is not quite as high as the baseline. Also, notice the trade-off between classification and localization that has occurred in the $\mathcal{L}_{IoU}$ samples.

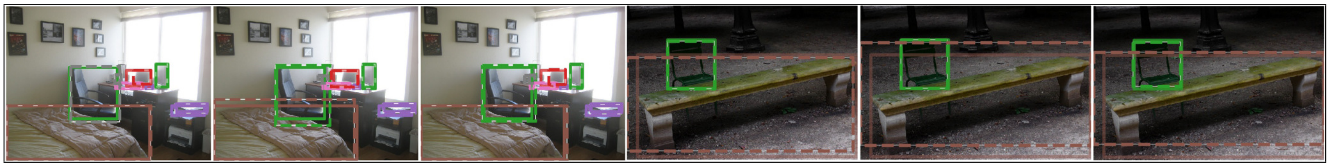

Figure 3. Example results from COCO validation using YOLO v3 [6] trained using (left to right) $\mathcal{L}_{GIoU}$, $\mathcal{L}_{IoU}$, and MSE losses. Ground truth is shown by a solid line and predictions are represented with dashed lines.

Fig. 4 shows two results from YOLO v3 [6] in cluttered scenes where $\mathcal{L}_{GIoU}$ performs much better on average in localization while maintaining good classification. Opacity of the dashed line for a prediction is set to the network output score. Solid white lines indicate ground truth.

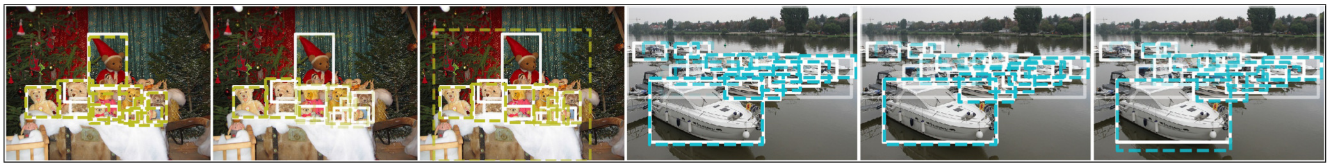

Figure 4. Example results from COCO validation using YOLO v3 [6] trained using (left to right) $\mathcal{L}_{GIoU}$, $\mathcal{L}_{IoU}$, and MSE losses. Ground truth is shown by a solid white line and predictions are represented with dashed lines. Opacity of the dashed line corresponds with the network's score for the given bounding box.

References

[1] G. Gilbert. Distance between sets. *Nature*, 239:174, 1972. 4

[2] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017. 9, 10

[3] S. Kosub. A note on the triangle inequality for the jaccard distance. *arXiv preprint arXiv:1612.02696*, 2016. 4

[4] M. Levandowsky and D. Winter. Distance between sets. *Nature*, 234(5323):34, 1971. 4

[5] A. H. Lipkus. A proof of the triangle inequality for the tanimoto distance. *Journal of Mathematical Chemistry*, 26(1-3):263–265, 1999. 4

[6] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018. 9, 10