

# Intersection over Union Bounds (Proposal)

Ashvin Srinivasan

Department of Information and Communication Engineering

Aalto University

Email: ashvin.1.srinivasan@aalto.fi

**Abstract**—We derive two theoretical robustness bounds for spectrogram segmentation in the context of wideband spectrum sensing, where small nuisance transformations, such as timing jitter or frequency drift, are modeled as smooth actions. Specifically, we establish: (i) a bound on the inflation of the expected per-pixel 0–1 classification risk, and (ii) a corresponding bound on the degradation of Intersection-over-Union (IoU), a key metric for signal localization. These results quantify how sensitive segmentation models are to realistic spectrum distortions and hold for standard DeepLab/UNet architectures trained with cross-entropy, without requiring any architectural changes. The bounds provide actionable insight into the stability of signal detection pipelines under physical perturbations common in dynamic RF environments.

## I. PROBLEM SETUP AND NOTATION

Let  $x \in \mathcal{X} \subset \mathbb{R}^{T \times F \times C}$  be a spectrogram-like input and  $y \in \{1, \dots, K\}^{T \times F}$  the pixelwise ground-truth labels. A segmentation network outputs logits  $f : \mathcal{X} \rightarrow \mathbb{R}^{K \times T \times F}$  with per-pixel prediction

$$\hat{y}(x)_p \triangleq \arg \max_{k \in \{1, \dots, K\}} f_k(x)_p,$$

for pixel  $p$  (indexing a time–frequency bin). We denote by  $\ell(\hat{y}(x)_p, y(p)) = \mathbf{1}\{\hat{y}(x)_p \neq y(p)\}$  the per-pixel 0–1 loss. The expression defines the per-pixel 0–1 loss, which evaluates to 1 if the predicted class label  $\hat{y}(x)_p$  at pixel  $p$  differs from the true class label  $y(p)$ , and 0 otherwise. This loss function is used to measure misclassification at the pixel level in a multi-class segmentation task, regardless of which specific class is involved. We define the expected pixel error rate (“risk”)

$$\mathcal{R} \triangleq \mathbb{E}[\ell(\hat{y}(x)_p, y(p))],$$

where the expectation is with respect to the data distribution and a uniformly sampled pixel. expectation is averaging over many images (or signals), and for each image, over a randomly selected pixel. In practice, this corresponds to computing the mean per-pixel error across the dataset.  $\mathbb{E} = \mathbb{E}_{x, y \sim P_{data}} \mathbb{E}_p \sim U(1, TF) = \mathbb{E}_{(x, y), p}$

A small nuisance acts on inputs,  $g \cdot x$ ,  $g \in G$ , with identity  $e \in G$ . Writing  $G$  (random variable) for a random nuisance, we are interested in the *risk inflation*

$$\Delta \mathcal{R} \triangleq \mathbb{E}[\ell(\hat{y}(G \cdot x)_p, y(p))] - \mathbb{E}[\ell(\hat{y}(x)_p, y(p))],$$

and in how this translates to a drop in Intersection-over-Union (IoU). A nuisance in this context refers to a transformation applied to the input  $x$  that does not alter its true label  $y(p)$ , but may cause a model to misclassify it. These transformations are

modeled by a smooth Lie group  $G$ , such as time or frequency shifts common in spectrum sensing. The action  $g \cdot x$ , for  $g \in G$ , represents applying such a transformation to the input, with  $e \in G$  denoting the identity (no change). The term  $\Delta \mathcal{R}$  quantifies risk inflation, the increase in expected pixel-wise 0–1 loss due to a random nuisance  $G$ . It measures the model’s sensitivity to benign, label-preserving variations. A small  $\Delta \mathcal{R}$  implies robustness, which is desirable in real-world settings like FR3 spectrum monitoring, where signals often experience minor but unavoidable distortions. Here  $G$  denotes a random transformation in the group  $\mathcal{G}$  (e.g., time shift, frequency drift, small time–frequency warp) acting on the input spectrogram via  $x \mapsto G \cdot x$ . We refer to  $G$  as a *nuisance* parameter because it changes the observed input  $G \cdot x$ , but does not change the ground-truth label at pixel  $p$  (the segmentation  $y(p)$ ). Moreover,  $G$  is not itself the quantity of interest: our goal is not to estimate  $G$ , but rather to obtain correct predictions  $f_k(G \cdot x)_p$  that are robust to such transformations. In the classical statistical sense,  $G$  is therefore a latent/random factor that affects the data and the logits  $f_k(G \cdot x)_p$ , but is not a parameter we wish to infer.

## II. ASSUMPTIONS

We now state explicit assumptions used in the propositions.

**Assumption 1** (Smooth small-deformation bound). *Let  $\mathcal{G}$  be a family of time–frequency transformations acting on input spectrograms via  $x \mapsto g \cdot x$ . There exists  $\rho > 0$  such that for each sample  $x$  there is a constant  $C_x > 0$  with*

$$\|g \cdot x - x\|_2 \leq C_x d_{\mathcal{G}}(e, g) \quad \text{for all } g \in \mathcal{G} \text{ with } d_{\mathcal{G}}(e, g) \leq \rho. \quad (1)$$

Assumption 1 formalizes the idea that small nuisance transformations of the time–frequency plane (such as slight time shifts, frequency offsets, or mild rescalings) induce only bounded changes in the input spectrogram. The set  $\mathcal{G}$  collects such transformations,  $e$  is the identity (no change), and  $d_{\mathcal{G}}(e, G)$  is a scalar that measures how large the transformation  $g$  is (for example, the Euclidean norm of its parameter vector). The assumption states that, whenever this magnitude  $d_{\mathcal{G}}(e, G)$  is small enough, the resulting change in the sampled input  $x$  (measured in Euclidean norm) is at most proportional to  $d_{\mathcal{G}}(e, G)$ , with a proportionality constant  $C_x$  that may depend on the particular sample. No additional differential-geometric structure is required; we only use that  $g \cdot x$  is a well-defined transformed version of  $x$  and that  $d_{\mathcal{G}}$  behaves as a reasonable

measure of transformation size. Assumption 1 models  $\mathcal{G}$  as a family of time–frequency transformations, i.e., a set of maps  $g : \mathcal{X} \rightarrow \mathcal{X}$  (typically forming a group with identity  $e$ ) acting on inputs via  $x \mapsto g \cdot x$ . For each sample  $x \in \mathcal{X}$  there exists a constant  $C_x > 0$  such that, for all “small” transformations with  $d_{\mathcal{G}}(e, g) \leq \rho$ ,  $\|g \cdot x - x\|_2 \leq C_x d_{\mathcal{G}}(e, g)$ . The factor  $d_{\mathcal{G}}(e, g)$  measures only the size of the transformation in the group, while  $C_x$  captures how sensitive the particular spectrogram  $x$  is to such perturbations. The universal radius  $\rho > 0$  defines a ball around the identity in transformation space where the action is locally Lipschitz; in Assumption 4 we require the random nuisance  $G$  to be supported inside this ball (ie,  $p(d_{\mathcal{G}}(e, G)) \leq \rho = 1$ ), so that  $d_{\mathcal{G}}(e, G) \leq \rho$  almost surely. The quantity  $d_{\mathcal{G}}(e, g)$  is a scalar measure of how large the transformation  $g \in \mathcal{G}$  is. If transformations are parameterized by a vector  $\theta(g) \in \mathbb{R}^m$  (e.g.,  $\theta(g) = (\Delta t, \Delta f)$  for time and frequency shifts), a natural choice is  $d_{\mathcal{G}}(e, g) = \|\theta(g)\|_2$  or  $d_{\mathcal{G}}(e, g) = |\Delta t| + |\Delta f|$ . More generally, if  $\mathcal{G}$  is a Lie group,  $d_{\mathcal{G}}$  could be any geodesic distance on the group. In this work we keep  $d_{\mathcal{G}}$  abstract and only require that it is nonnegative, equals zero at the identity  $e$ , and that Assumption 1’s Lipschitz bound holds in terms of this scalar. Intuitively,  $d_{\mathcal{G}}(e, g)$  answers: “how big is this time/frequency distortion?”, measured in whatever normalized units (e.g., seconds and hertz, or their Euclidean norm) are appropriate for the chosen parameterization.

**Assumption 2** (Logit-Lipschitz network). *The logit map  $f : \mathcal{X} \rightarrow \mathbb{R}^{K \times T \times F}$  is globally  $L$ -Lipschitz [2]:*

$$\|f(x') - f(x)\|_2 \leq L \|x' - x\|_2, \quad \forall x, x' \in \mathcal{X}. \quad (2)$$

In Assumption 2 we use a single Lipschitz constant  $L$  for the logit map because this is a property we can enforce by design (e.g. spectral normalization / Parseval networks), and standard results guarantee a global Lipschitz bound for the network. In contrast, Assumption 1 describes the physical action  $g \cdot x$ . For a smooth action we only get a local Lipschitz constant around the identity for each fixed sample  $x$ ; there is no reason to expect a uniform constant over all  $x$  on a non-compact data space. To keep the nuisance model realistic and as weak as possible, we allow a per-sample constant  $C_x$  and only assume  $E[C_x] < \infty$ . A global constant would just be the special case  $C_x = C$  but that would be a stronger condition and often not realistic. If the group action admits a uniform bound  $\|g \cdot x - x\|_2 \leq C_x d_{\mathcal{G}}(e, G)$  then all results below hold with  $E[C_x]$  replaced by  $C$ , we have more general per-sample form to avoid imposing an unnecessary global bound.

**Assumption 3** (Margin). *For pixel  $p$  with true class  $y(p)$ , define the margin*

$$\gamma_p(x) \triangleq f_{y(p)}(x)_p - \max_{k \neq y(p)} f_k(x)_p. \quad (3)$$

*For a chosen threshold  $\gamma_0 > 0$ , let  $\mathcal{M}_{\gamma_0}$  be the event  $\{\gamma_p(x) \geq \gamma_0\}$  and*

$$\eta(\gamma_0) \triangleq \mathbb{P}(\neg \mathcal{M}_{\gamma_0}) \quad (4)$$

*the fraction of low-margin pixels.*

Assumption 3 introduces the concept of margin at each pixel, defined as the difference between the logit assigned to the true class and the highest logit assigned to any incorrect class. This margin, denoted  $\gamma_p(x)$ , reflects the model’s confidence in its prediction at pixel  $p$ . A large positive margin implies high confidence and robustness to small perturbations, while a negative margin indicates misclassification. By choosing a threshold  $\gamma_0 > 0$ , we define the event that a pixel has sufficient margin, and the quantity  $\eta(\gamma_0)$  captures the fraction of pixels that are either already misclassified or are close to decision boundaries (i.e., low-margin). This assumption enables the separation of vulnerable pixels from confident ones, which is crucial for bounding the model’s sensitivity to input distortions.

**Assumption 4** (Nuisance statistics). *The random nuisance  $G$  is independent of  $(x, y)$  and supported within the small-deformation regime of (1). Let*

$$\mu_1 \triangleq \mathbb{E}[d_{\mathcal{G}}(e, G)] < \infty. \quad (5)$$

*Optionally, we will also consider a sub-Gaussian tail:  $\mathbb{P}\{d_{\mathcal{G}}(e, G) \geq t\} \leq \exp(-t^2/(2\sigma^2))$  for some  $\sigma > 0$  [3].*

Assumption 4 models the nuisance transformation  $G$  as a random variable that is independent of the data sample  $(x, y)$  and constrained to the small-deformation regime. This captures the realistic scenario in spectrum sensing where input perturbations, such as timing offsets, frequency drifts, or oscillator phase noise, are small, stochastic, and unrelated to the signal class. The assumption that  $G$  has bounded mean deformation ensures that the average size of such perturbations is finite, which is reasonable given modern synchronization and calibration in RF front-ends. Additionally, assuming a sub-Gaussian tail for  $d_{\mathcal{G}}(e, G)$  reflects the empirical observation that such physical impairments exhibit rapidly decaying probability for large deviations, due to hardware constraints. Sub-Gaussianity enables exponentially decaying bounds on risk inflation, making it a reasonable assumption for analyzing the robustness of classification models under realistic spectrum distortions.

### III. LEMMAS

**Lemma 1** (Logit change under nuisance). *Under Assumptions 1 and 2, for any pixel  $p$  and class  $k$ ,*

$$|f_k(G \cdot x)_p - f_k(x)_p| \leq L C_x d_{\mathcal{G}}(e, G). \quad (6)$$

*Proof.* Apply (2) with  $x' = G \cdot x$  and  $x$ , then apply (1):

$$\|f(G \cdot x) - f(x)\|_2 \leq L \|G \cdot x - x\|_2 \leq L C_x d_{\mathcal{G}}(e, G).$$

The bound holds componentwise in particular for the scalar entry  $(k, p)$ .  $\square$

**Lemma 2** (Margin stability implies label stability). *Let  $p$  be a pixel with true class  $y \equiv y(p)$  and logits  $f_k(x)_p$ . Define the logit margin [1]*

$$\gamma_p(x) \triangleq f_y(x)_p - \max_{k \neq y} f_k(x)_p. \quad (7)$$

If  $\gamma_p(x) \geq \gamma_0 > 0$  and  $2LC_x d_G(e, G) < \gamma_0$ , then the predicted label at  $p$  is unchanged by the nuisance:

$$\arg \max_k f_k(G \cdot x)_p = \arg \max_k f_k(x)_p = y.$$

*Proof.* By the definition in (7),

$$f_y(x)_p > \max_{k \neq y} f_k(x)_p \Rightarrow \arg \max_k f_k(x)_p = y. \quad (8)$$

Let  $d \triangleq d_G(e, G)$ . By Lemma 1, for every class  $k$ ,

$$|f_k(G \cdot x)_p - f_k(x)_p| \leq LC_x d. \quad (9)$$

Applying (9) with  $k = y$  yields a lower bound on the true-class logit under the nuisance:

$$f_y(G \cdot x)_p \geq f_y(x)_p - LC_x d. \quad (10)$$

For any  $k \neq y$ , (9) gives  $f_k(G \cdot x)_p \leq f_k(x)_p + LC_x d$ ; maximizing over  $k \neq y$  then implies

$$\max_{k \neq y} f_k(G \cdot x)_p \leq \max_{k \neq y} f_k(x)_p + LC_x d. \quad (11)$$

Subtracting (11) from (10) compares the post-nuisance true logit to the strongest competitor:

$$\begin{aligned} & f_y(G \cdot x)_p - \max_{k \neq y} f_k(G \cdot x)_p \\ & \geq (f_y(x)_p - LC_x d) - (\max_{k \neq y} f_k(x)_p + LC_x d) \\ & = f_y(x)_p - \max_{k \neq y} f_k(x)_p - 2LC_x d. \end{aligned} \quad (12)$$

$\underbrace{\hspace{10em}}_{= \gamma_p(x)}$

Under the small-nuisance condition  $2LC_x d < \gamma_0 \leq \gamma_p(x)$ , the right-hand side of (12) is strictly positive, hence

$$f_y(G \cdot x)_p > \max_{k \neq y} f_k(G \cdot x)_p,$$

which shows that  $y$  still attains the argmax after applying  $G$  and therefore  $\arg \max_k f_k(G \cdot x)_p = y$ .

Finally, we note that the identity of the runner-up class (i.e., the maximizer over  $\{f_k(\cdot)_p : k \neq y\}$ ) may change between  $x$  and  $G \cdot x$ . Using the maximum in (11) covers this worst case, so the conclusion holds regardless of which competitor is strongest.  $\square$

#### IV. RISK INFLATION BOUND

**Proposition 1** (Risk inflation under small nuisances). *Under Assumptions 1–4, for any  $\gamma_0 > 0$ ,*

$$\Delta \mathcal{R} \leq \eta(\gamma_0) + \frac{2L \mathbb{E}[C_x] \mu_1}{\gamma_0}. \quad (13)$$

*If, additionally,  $d_G(e, G)$  is sub-Gaussian with parameter  $\sigma$ , then*

$$\Delta \mathcal{R} \leq \eta(\gamma_0) + \mathbb{E} \left[ \exp \left( -\frac{\gamma_0^2}{8L^2 C_x^2 \sigma^2} \right) \right]. \quad (14)$$

*Proof.* We bound the excess error probability per pixel; averaging over pixels and data yields  $\Delta \mathcal{R}$ .

Let

$$A \triangleq \{\hat{y}(G \cdot x)_p \neq y(p)\}, \quad B \triangleq \mathcal{M}_{\gamma_0} = \{\gamma_p(x) \geq \gamma_0\}.$$

Since  $\Omega = B \cup \neg B$  is a disjoint partition, we have the exact decomposition

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \neg B). \quad (15)$$

On the low-margin set  $\neg B$ , we simply upper bound

$$\mathbb{P}(A \cap \neg B) \leq \mathbb{P}(\neg B) = \eta(\gamma_0). \quad (16)$$

Combining (15) and (16) yields the relaxed inequality used in the sequel:

$$\mathbb{P}\{\hat{y}(G \cdot x)_p \neq y(p)\} \leq \eta(\gamma_0) + \mathbb{P}(\hat{y}(G \cdot x)_p \neq y(p) \wedge B). \quad (17)$$

Next, define also

$$D \triangleq \left\{ d_G(e, G) \geq \frac{\gamma_0}{2LC_x} \right\}.$$

By Lemma 2, on the good-margin set  $B$  a label flip cannot happen unless the nuisance is large enough to eat a margin  $\gamma_0$ . Formally,

$$B \wedge \neg D \Rightarrow \neg A \iff B \wedge A \Rightarrow D, \quad (18)$$

i.e., **If a pixel is both well-margin and misclassified, then the nuisance must have been large.** In set notation, (18) yields the inclusions

$$A \wedge B \subseteq D. \quad (19)$$

Applying the monotonicity of probability to (19) gives

$$\begin{aligned} \mathbb{P}(\hat{y}(G \cdot x)_p \neq y(p) \wedge \mathcal{M}_{\gamma_0}) &= \mathbb{P}(A \wedge B) \\ &\leq \mathbb{P}(D) \\ &= \mathbb{P}\left(d_G(e, G) \geq \frac{\gamma_0}{2LC_x}\right). \end{aligned} \quad (20)$$

*Please Note.*  $D$  is only a *necessary* condition for a flip on  $B$  (large nuisance); it contains outcomes with both correct and incorrect predictions. Independence of  $G$  from  $(x, y)$  is *not* needed for (20); it is used later when we bound  $\mathbb{P}(D)$  (e.g., via Markov or sub-Gaussian tails).

Conditioning on  $x$  and using independence (Assumption 4), apply Markov's inequality to the nonnegative random variable  $d_G(e, G)$ :

$$\mathbb{P}\left(d_G(e, G) \geq \frac{\gamma_0}{2LC_x} \mid x\right) \leq \frac{2LC_x}{\gamma_0} \mathbb{E}[d_G(e, G)] = \frac{2LC_x}{\gamma_0} \mu_1.$$

Taking expectation over  $x$  and pixels gives

$$\mathbb{P}(\hat{y}(G \cdot x)_p \neq y(p) \wedge \mathcal{M}_{\gamma_0}) \leq \frac{2L \mathbb{E}[C_x] \mu_1}{\gamma_0}.$$

Adding the low-margin mass  $\eta(\gamma_0)$  yields (13).

If in addition  $d_G(e, G)$  is sub-Gaussian with parameter  $\sigma$ , then for any  $a > 0$ ,

$$\mathbb{P}(d_G(e, G) \geq a) \leq \exp\left(-\frac{a^2}{2\sigma^2}\right).$$

Setting  $a = \gamma_0/(2LC_x)$  in (20) yields

$$\mathbb{P}(\hat{y}(G \cdot x)_p \neq y(p) \wedge \mathcal{M}_{\gamma_0}) \leq \exp\left(-\frac{\gamma_0^2}{8L^2 C_x^2 \sigma^2}\right).$$

Averaging over  $x$  and pixels and adding  $\eta(\gamma_0)$  gives (14).  $\square$

## V. FROM PIXEL RISK TO IOU

We relate pixel error to IoU for a given foreground class  $k$ .

**Definition 1** (Per-class IoU). Let  $S \subset \{1, \dots, T\} \times \{1, \dots, F\}$  be the set of foreground pixels for class  $k$ , with  $A = |S|$  and total pixels  $N = TF$ . Let (True Positives) TP, (False Positives) FP, (False Negatives) FN be the usual counts and define  $E = FP + FN$ . The Intersection over Union (IoU) is

$$\text{IoU}_k = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}.$$

**Lemma 3** (IoU lower bound via total pixel error). Let  $A = |S|$  be the number of foreground (GT) pixels for class  $k$ , and let TP, FP, FN be the usual counts with  $E = FP + FN$ . Then

$$\text{IoU}_k = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} = \frac{A - \text{FN}}{A + \text{FP}} \geq \frac{A - E}{A + E}. \quad (21)$$

Moreover, if  $E \leq A$ , then

$$\text{IoU}_k \geq \frac{A - E}{A + E} = 1 - \frac{2E}{A + E} \geq 1 - \frac{2E}{A}. \quad (22)$$

In normalized form  $e = E/N$  and  $\alpha = A/N$ ,

$$\text{IoU}_k \geq \frac{\alpha - e}{\alpha + e} \geq 1 - \frac{2e}{\alpha} \quad (\text{if } e \leq \alpha). \quad (23)$$

*Proof.* First, since  $A = \text{TP} + \text{FN}$ , we have  $\text{TP} = A - \text{FN}$  and

$$\text{TP} + \text{FP} + \text{FN} = (A - \text{FN}) + \text{FP} + \text{FN} = A + \text{FP},$$

which gives the identity in (21).

For the bound, observe that  $\text{FP} \geq 0$  implies

$$A - \text{FN} \geq A - (\text{FN} + \text{FP}) = A - E,$$

and  $\text{FN} \geq 0$  implies

$$A + \text{FP} \leq A + (\text{FP} + \text{FN}) = A + E.$$

With a larger numerator and a smaller denominator, the fraction increases, hence

$$\frac{A - \text{FN}}{A + \text{FP}} \geq \frac{A - E}{A + E},$$

proving the inequality in (21).

For (22), note that

$$\frac{A - E}{A + E} = 1 - \frac{2E}{A + E}.$$

If additionally  $E \leq A$ , then  $A + E \geq A$  so  $\frac{1}{A+E} \leq \frac{1}{A}$  and

$$1 - \frac{2E}{A + E} \geq 1 - \frac{2E}{A}.$$

Dividing numerator and denominator by  $N$  yields (23).  $\square$

**Remark 1.** The exact identity is  $\text{IoU}_k = (A - \text{FN}) / (A + \text{FP})$ . The bound  $\text{IoU}_k \geq (A - E) / (A + E)$  is conservative but convenient because it depends only on the total pixel error  $E = \text{FP} + \text{FN}$ .

**Proposition 2** (IoU-drop bound). Let  $e_0$  be the baseline pixel error rate for class  $k$  and  $e_1 = e_0 + \Delta \mathcal{R}_k$  the error rate after the nuisance. Define  $\alpha_k = A/N$  for class  $k$ . Then

$$\text{IoU}_k^{\text{after}} \geq \frac{\alpha_k - e_1}{\alpha_k + e_1}, \quad \text{IoU}_k^{\text{before}} \geq \frac{\alpha_k - e_0}{\alpha_k + e_0}. \quad (24)$$

Hence the IoU drop  $\Delta \text{IoU}_k \triangleq \text{IoU}_k^{\text{before}} - \text{IoU}_k^{\text{after}}$  satisfies

$$\Delta \text{IoU}_k \leq \frac{2\alpha_k}{(\alpha_k + e_0)^2} \Delta \mathcal{R}_k. \quad (25)$$

*Proof.* The lower bounds in (24) are direct from Lemma 3. Define  $f(e) = \frac{\alpha_k - e}{\alpha_k + e}$  for  $e \in [0, \alpha_k]$ . Then  $f$  is differentiable with

$$f'(e) = -\frac{2\alpha_k}{(\alpha_k + e)^2} \quad \text{and} \quad f \text{ is decreasing in } e.$$

By the mean value theorem, for some  $\xi \in [e_0, e_1]$ ,

$$f(e_0) - f(e_1) = -f'(\xi)(e_1 - e_0) = \frac{2\alpha_k}{(\alpha_k + \xi)^2} \Delta \mathcal{R}_k \leq \frac{2\alpha_k}{(\alpha_k + e_0)^2} \Delta \mathcal{R}_k,$$

since  $(\alpha_k + \xi) \geq (\alpha_k + e_0)$ . This is (25).  $\square$

**Corollary 1** (Combined action,  $g \in G \Rightarrow \text{IoU}$  bound). Combining propos 1 (Markov form) with propos 2 gives, for any  $\gamma_0 > 0$ ,

$$\Delta \text{IoU}_k \leq \frac{2\alpha_k}{(\alpha_k + e_0)^2} \left( \eta_k(\gamma_0) + \frac{2L \mathbb{E}[C_x] \mu_1}{\gamma_0} \right). \quad (26)$$

Under the sub-Gaussian tail, we similarly obtain

$$\Delta \text{IoU}_k \leq \frac{2\alpha_k}{(\alpha_k + e_0)^2} \left( \eta_k(\gamma_0) + \mathbb{E} \left[ e^{-\frac{\gamma_0^2}{8L^2 C_x^2 \sigma^2}} \right] \right). \quad (27)$$

Averaging over classes yields the mIoU version.

## REFERENCES

- [1] Yusuke Tsuzuku, Issei Sato, and Masashi Sugiyama. 2018. Lipschitz-margin training: scalable certification of perturbation invariance for deep neural networks. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). Curran Associates Inc., Red Hook, NY, USA, 6542–6551.
- [2] Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, and Nicolas Usunier. 2017. Parseval networks: improving robustness to adversarial examples. In Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17). JMLR.org, 854–863.
- [3] <https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-2.pdf>
- [4] <https://arxiv.org/pdf/math/0305427#:~:text=but%20quite%20useful%20result%20which,For%20a%20di%EF%AC%80erentiable>