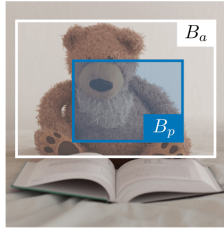
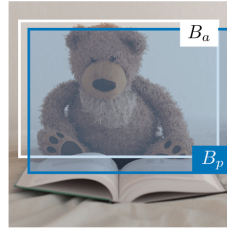


$$\text{IoU}(B_p, B_a) = 0.1$$



$$\text{IoU}(B_p, B_a) = 0.5$$



$$\text{IoU}(B_p, B_a) = 0.9$$

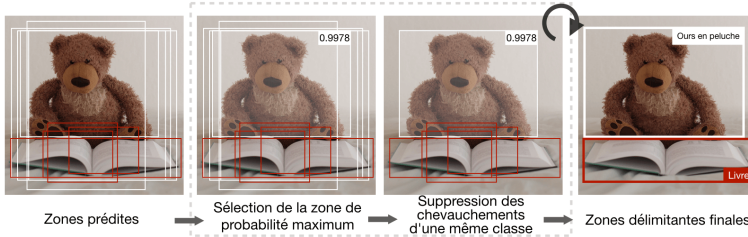
Remarque : on a toujours $\text{IoU} \in [0,1]$. Par convention, la prédiction B_p d'une zone délimitante est considérée comme étant satisfaisante si l'on a $\text{IoU}(B_p, B_a) \geq 0.5$.

□ **Zone d'accroche** – La technique des zones d'accroche (en anglais *anchor boxing*) sert à prédire des zones délimitantes qui se chevauchent. En pratique, on permet au réseau de prédire plus d'une zone délimitante simultanément, où chaque zone prédite doit respecter une forme géométrique particulière. Par exemple, la première prédiction peut potentiellement être une zone rectangulaire d'une forme donnée, tandis qu'une seconde prédiction doit être une zone rectangulaire d'une autre forme.

□ **Suppression non-max** – La technique de suppression non-max (en anglais *non-max suppression*) a pour but d'enlever des zones délimitantes qui se chevauchent et qui prédisent un seul et même objet, en sélectionnant les zones les plus représentatives. Après avoir enlevé toutes les zones ayant une probabilité prédite de moins de 0.6, les étapes suivantes sont répétées pour éliminer les zones redondantes :

Pour une classe donnée,

- Étape 1 : Choisir la zone ayant la plus grande probabilité de prédiction.
- Étape 2 : Enlever toute zone ayant $\text{IoU} \geq 0.5$ avec la zone choisie précédemment.



□ **YOLO** – L'algorithme You Only Look Once (YOLO) est un algorithme de détection d'objet qui fonctionne de la manière suivante :

- Étape 1 : Diviser l'image d'entrée en une grille de taille $G \times G$.
- Étape 2 : Pour chaque cellule, faire tourner un CNN qui prédit y de la forme suivante :

$$y = \left[\underbrace{p_c, b_x, b_y, b_h, b_w, c_1, c_2, \dots, c_p}_{\text{répété } k \text{ fois}} \right]^T \in \mathbb{R}^{G \times G \times k \times (5+p)}$$

où p_c est la probabilité de détecter un objet, b_x, b_y, b_h, b_w sont les propriétés de la zone délimitante détectée, c_1, \dots, c_p est une représentation binaire (en anglais *one-hot representation*) de l'une des p classes détectée, et k est le nombre de zones d'accroche.

- Étape 3 : Faire tourner l'algorithme de suppression non-max pour enlever des doublons potentiels qui chevauchent des zones délimitantes.