

# Recherche d'informations visuelles

Nicolas Thome

Université Pierre et Marie Curie (UPMC)  
Master Informatique - Spécialité DAC  
Laboratoire d'Informatique de Paris 6 (LIP6)



# Plan

## 1 Apprentissage Interactif pour la RI

- Contexte
- Solutions

## 2 Apprentissage faiblement supervisé

## 3 Apprentissage avec des données "exotiques"

# Apprentissage Interactif

## Contexte

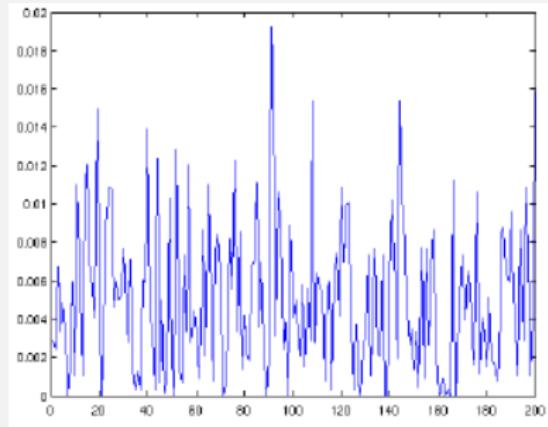
- ① Sélection d'une image requête par l'utilisateur
- ② Tri de la base par ordre décroissant de similarité
- ③ Annotations d'images  $\oplus/\ominus$  par l'utilisateur
  - $\Rightarrow$  Quelles images proposer à annoter ?

# Apprentissage Interactif

## Contexte



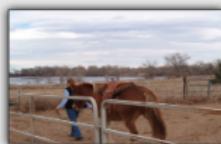
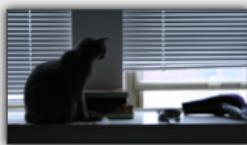
Image requête



Décrise dans un espace de représentation (e.g. BoW)

# Apprentissage Interactif

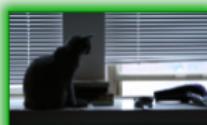
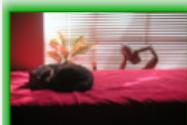
## Tri de la base par ordre décroissant de similarité / requête



# Apprentissage Interactif

## Similarité : limitations

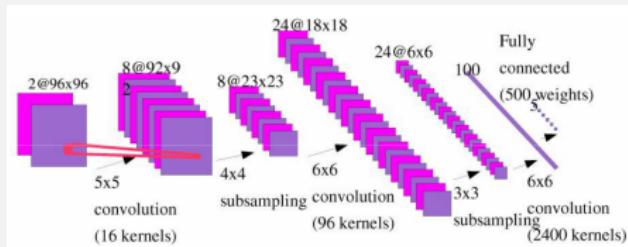
- Problème du fossé sémantique
- Solution : annotations et apprentissage supervisé



# Apprentissage Interactif

## Annotations et apprentissage supervisé

- Modèles linéaires :  $f(x) = \langle w; \phi(x) \rangle (+b)$ 
  - perceptron, LDA, SVM, etc
- Modèles non linéaires :  $f(x)$  non linéaire
  - Noyaux (kernel trick) :
    - $w = \sum_i \alpha_i y_i \phi(x_i)$
    - $K(x, x_i) = \langle \phi(x); \phi(x_i) \rangle$
    - $f(x) = \sum_i \alpha_i y_i K(x, x_i) (+b)$
- Modèles profonds vs flat



# Apprentissage Interactif

Annotations  $\Rightarrow$  apprentissage d'une fonction de décision  $f(x)$



Frontière de  
décision  $f(x)$



# Apprentissage Interactif

## Tri de la base par ordre décroissant / $f(x)$

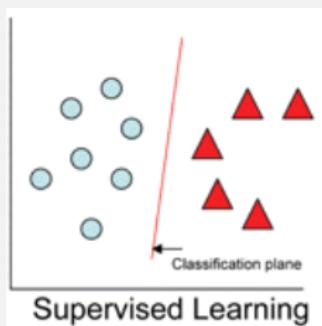
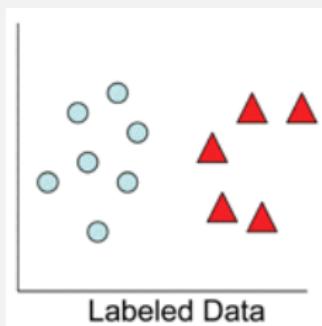


Et on recommence une session d'annotation ...

# Apprentissage Interactif

## Apprentissage supervisé

- Apprentissage supervisé classique :
  - Un ensemble d'apprentissage
    - Entrainement d'un classifieur sur cet ensemble
  - Evaluation des performances sur le reste de la base (test)
  - Les exemples sont supposés *i.i.d* : Indépendants et identiquement distribués



# Apprentissage Interactif

## Apprentissage interactif et actif

- Base non annotée (au départ)
- Interaction avec l'utilisateur pour effectuer une annotation itérative
- La catégorie sémantique recherchée peut varier
  - D'un utilisateur à l'autre
  - D'une session à l'autre
- Objectif de l'annotation : améliorer la qualité de la recherche sémantique
- **Question centrale :** Quelles (nouvelles) images présenter à l'utilisateur pour annotation ?
  - Comment choisir l'ensemble d'apprentissage pour optimiser :
    - La qualité du modèle appris
    - La vitesse de convergence de l'algorithme

# Plan

## 1 Apprentissage Interactif pour la RI

- Contexte
- Solutions

## 2 Apprentissage faiblement supervisé

## 3 Apprentissage avec des données "exotiques"

# Apprentissage Interactif

## Quelles images annoter ?

- Tirage aléatoire d'images dans la bases (données i.i.d) ?
- avoir des heuristiques
- Images au top du classement
- Images incertaines
- Diversité ?

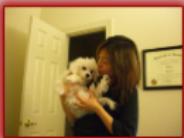
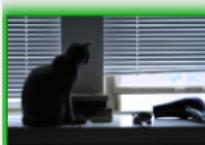
# Quelles images annoter ?

## Images aléatoires

- Tirage aléatoire d'images dans la bases
- Convergence lente car n'exploite pas le classifieur appris de manière incrémentale
- Suppose implicitement que les données sont i.i.d
- A rapprocher d'un apprentissage supervisé classique

# Quelles images annoter ?

## Images au top du classement



# Quelles images annoter ?

## Images incertaines

- Alternative aux images du top : quelles autres images ?
  - Moins bien classées, mais dont l'annotation aiderait "au mieux" l'apprentissage d'une bonne fonction de classification
- Annoter les images les plus "incertaines" [TK02]
  - Les images proches de la frontière de décision,  $|f(x)|$  proche de 0
  - Stratégie "optimale" ([TK02]) : Coupe l'espace de recherche de manière à converger asymptotiquement "au mieux" (au plus vite) vers le classifieur appris avec l'ensemble des données



# Quelles images annoter ?

## Combinaison Top + images incertaines

- Classifieur fiable (frontière de décision pertinente) : annoter les incertaines



- Classifieur pas fiable (au début) : frontière de décision très bruitée
  - Mieux vaut annoter le top



# Quelles images annoter ?

## Diversité

- Ensemble des données  $\mathcal{D}$ ,  $\mathcal{A}$  annotations,  $\mathcal{U} = \mathcal{D} \setminus \mathcal{A}$  le reste
- Images incertaines : images pour lesquelles  $|f(x)|$  proche de 0 :

$$j^* = \arg \min_{x_j \in \mathcal{U}} (|f_{\mathcal{A}}(x_j)|)$$

- Critère de diversité : [Bri03]

$$j^* = \arg \min_{x_j \in \mathcal{U}} \left[ \max_{x_i \in \mathcal{A}} \frac{|\langle \phi(x_i); \phi(x_j) \rangle|}{\|\phi(x_i)\| \cdot \|\phi(x_j)\|} \right] = \arg \min_{x_j \in \mathcal{U}} \left[ \max_{x_i \in \mathcal{A}} \frac{|K(x_i, x_j)|}{\sqrt{K(x_i, x_i)K(x_j, x_j)}} \right] \quad (1)$$

- $K(x_i, x_i) = \langle \phi(x_i); \phi(x_i) \rangle$  : alignement (angle)
- Image sélectionnée "différente" de toutes les images déjà annotées

# Quelles images annoter ?

## Incertaines + Diversité

- Diversité :

$$\begin{aligned}
 j^* &= \arg \min_{x_j \in \mathcal{U}} \left[ \max_{x_i \in \mathcal{A}} \frac{|\langle \phi(x_i); \phi(x_j) \rangle|}{\|\phi(x_i)\| \cdot \|\phi(x_j)\|} \right] = \arg \min_{x_j \in \mathcal{U}} \left[ \max_{x_i \in \mathcal{A}} \frac{|K(x_i, x_j)|}{\sqrt{K(x_i, x_i)K(x_j, x_j)}} \right] \\
 &= \arg \min_{x_j \in \mathcal{U}} \left[ \max_{x_i \in \mathcal{A}} K^*(x_i, x_j) \right]
 \end{aligned} \tag{2}$$

- Incertaines + diversité :

$$j^* = \arg \min_{x_j \in \mathcal{U}} \left[ \lambda |f_{\mathcal{A}}(x_j)| + (1 - \lambda) \max_{x_i \in \mathcal{A}} K^*(x_i, x_j) \right]$$

# Plan

1 Apprentissage Interactif pour la RI

2 Apprentissage faiblement supervisé

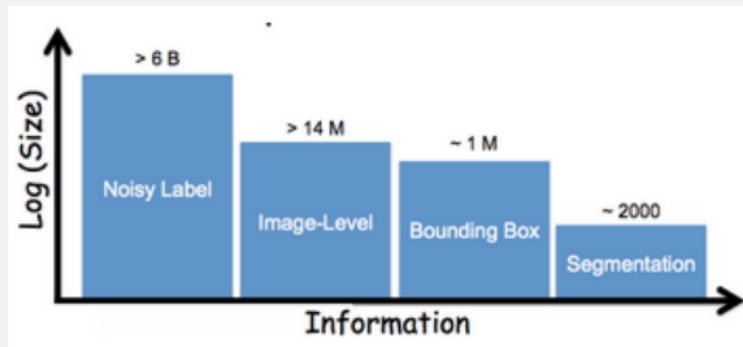
- Contexte
- Latent Structural SVM (LSSVM)
- Exemples d'applications en vision

3 Apprentissage avec des données "exotiques"

# Apprentissage faiblement supervisé

## Motivation

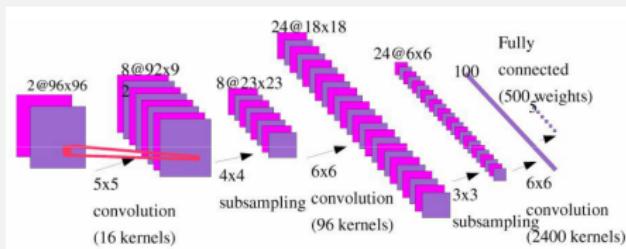
- Big data : beaucoup de données
- MAIS peu d'annotations en proportion
  - : surtout des annotations fines/précises
- $\Rightarrow$  Apprendre avec des annotations faibles



# Apprentissage faiblement supervisé

## Positionnement / deep learning

- Réseaux profond (ConvNets) : état de l'art pour tâches de classification large échelle
  - Features deep très bonnes en transfert learning
- MAIS ConvNets : structure assez figée
  - Apprendre des invariances fortes difficile
- ⇒ Combiner features deep et méthodes d'apprentissage faiblement supervisées



# Plan

1 Apprentissage Interactif pour la RI

2 Apprentissage faiblement supervisé

- Contexte
- Latent Structural SVM (LSSVM)
- Exemples d'applications en vision

3 Apprentissage avec des données "exotiques"

# Apprentissage faiblement supervisé & structuré

## Étude d'un formalisme

- Latent Structural SVM [YJ09]
- la relation entrée-sortie ne peut pas être complètement caractériser par l'ensemble d'apprentissage  $\mathcal{S} = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)\} \in (\mathcal{X} \times \mathcal{Y})^n$ , mais dépend d'un ensemble de variables cachées  $\mathbf{h} \in \mathcal{H}$
- évaluer un Latent Structural SVM revient à apprendre une prédiction de la forme :

$$(\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i; \mathbf{w})) = \underset{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}}{\operatorname{argmax}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle$$

- $\Psi(\mathbf{x}, \mathbf{y}, \mathbf{h})$  décrit la relation entre l'entrée  $\mathbf{x}$ , la sortie structurée  $\mathbf{y}$  et les variables cachées  $\mathbf{h}$

# Latent Structural SVM

## Extension de la fonction de coût

- mesure la différence entre 2 paires

$$\Delta((\mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i, \mathbf{w})), (\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})))$$

avec  $(\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i; \mathbf{w})) = \underset{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}}{\operatorname{argmax}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle$

et  $\mathbf{h}^*(\mathbf{x}_i; \mathbf{w}) = \underset{\mathbf{h} \in \mathcal{H}}{\operatorname{argmax}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle$

# Latent Structural SVM

majoration de la fonction de coût (cas général) :

$$\begin{aligned}
 & \Delta((\mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i, \mathbf{w})), (\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w}))) \\
 & \leq \Delta((\mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i, \mathbf{w})), (\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w}))) \\
 & \quad + \underbrace{\langle \mathbf{w}, \Psi(\mathbf{x}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) \rangle - \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i, \mathbf{w})) \rangle}_{\geq 0} \\
 & \leq \left( \max_{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle \right) + \Delta((\mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i, \mathbf{w})), (\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w}))) \\
 & \quad - \left( \max_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle \right)
 \end{aligned}$$

# Latent Structural SVM

- **hypothèse** : la fonction de coût ne dépend pas de pas de la variable cachée  $\mathbf{h}^*(\mathbf{x}_i, \mathbf{w})$

$$\Delta((\mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i, \mathbf{w})), (\hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w}))) = \Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w}))$$

- nouvelle majoration :

$$\begin{aligned}\Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) &\leq \left( \max_{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle \right) \\ &+ \Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) - \left( \max_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle \right)\end{aligned}$$

# Problème d'optimisation du Latent Structural SVM

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^n \left( \max_{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle + \Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) \right) - \frac{C}{n} \sum_{i=1}^n \left( \max_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle \right)$$

# Problème d'optimisation du Latent Structural SVM

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^n \left( \max_{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle + \Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) \right)$$

$$- \frac{C}{n} \sum_{i=1}^n \left( \max_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle \right)$$

- différence de 2 fonctions convexes :  $f(\mathbf{w}) - g(\mathbf{w})$

$$f(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^n \left( \max_{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle + \Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) \right)$$

$$g(\mathbf{w}) = \frac{C}{n} \sum_{i=1}^n \left( \max_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle \right)$$

# Problème d'optimisation du Latent Structural SVM

---

## Algorithm 1 Concave-Convex Procedure (CCCP)

---

- 1: Set  $t = 0$  and initialize  $\mathbf{w}_0$
  - 2: **repeat**
  - 3:   Find hyperplane  $\mathbf{v}_t$  such that  $-g(\mathbf{w}) \leq -g(\mathbf{w}_t) + (\mathbf{w} - \mathbf{w}_t) \cdot \mathbf{v}_t$  for all  $\mathbf{w}$
  - 4:   Solve  $\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \mathbf{w} \cdot \mathbf{v}_t$
  - 5:   Set  $t = t + 1$
  - 6: **until**  $[f(\mathbf{w}_t) - g(\mathbf{w}_t)] - [f(\mathbf{w}_{t-1}) - g(\mathbf{w}_{t-1})] < \epsilon$
- 

- Rédoudre le problème écrit comme une différence de fonctions convexes : Concave-Convex Procedure [YR03] (CCCP)
  - ⊕ Garanties de convergence (vers un minimum local)
  - ⊕ Critère formel d'arrêt de l'algorithme

# Problème d'optimisation du Latent Structural SVM

---

## Algorithm 1 Concave-Convex Procedure (CCCP)

---

- 1: Set  $t = 0$  and initialize  $\mathbf{w}_0$
  - 2: **repeat**
  - 3:   Find hyperplane  $\mathbf{v}_t$  such that  $-g(\mathbf{w}) \leq -g(\mathbf{w}_t) + (\mathbf{w} - \mathbf{w}_t) \cdot \mathbf{v}_t$  for all  $\mathbf{w}$
  - 4:   Solve  $\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \mathbf{w} \cdot \mathbf{v}_t$
  - 5:   Set  $t = t + 1$
  - 6: **until**  $[f(\mathbf{w}_t) - g(\mathbf{w}_t)] - [f(\mathbf{w}_{t-1}) - g(\mathbf{w}_{t-1})] < \epsilon$
- 

$$3 \rightarrow \forall i, \quad \mathbf{h}^*(\mathbf{x}_i; \mathbf{w}) = \operatorname{argmax}_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}) \rangle$$

$$\mathbf{v}_t = \sum_{i=1}^n \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i; \mathbf{w}))$$

# Problème d'optimisation du Latent Structural SVM

---

## Algorithm 1 Concave-Convex Procedure (CCCP)

---

- 1: Set  $t = 0$  and initialize  $\mathbf{w}_0$
  - 2: **repeat**
  - 3:   Find hyperplane  $\mathbf{v}_t$  such that  $-g(\mathbf{w}) \leq -g(\mathbf{w}_t) + (\mathbf{w} - \mathbf{w}_t) \cdot \mathbf{v}_t$  for all  $\mathbf{w}$
  - 4:   Solve  $\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \mathbf{w} \cdot \mathbf{v}_t$
  - 5:   Set  $t = t + 1$
  - 6: **until**  $[f(\mathbf{w}_t) - g(\mathbf{w}_t)] - [f(\mathbf{w}_{t-1}) - g(\mathbf{w}_{t-1})] < \epsilon$
- 

3 : construit un hyperplan qui majore la fonction concave  $g$   
⇒ l'optimisation de la ligne 4 est convexe

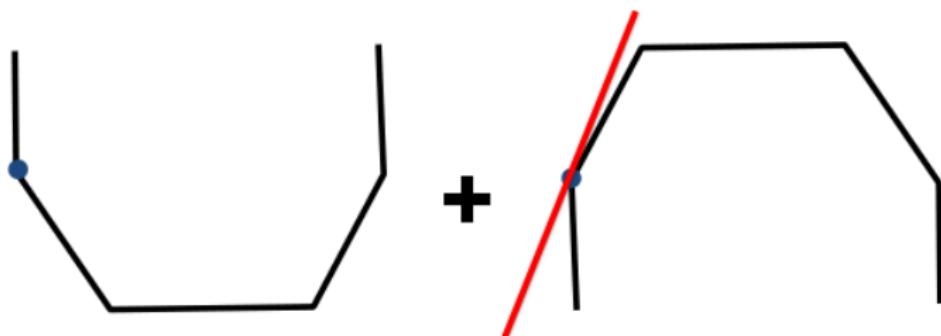
# Problème d'optimisation du Latent Structural SVM

---

## Algorithm 1 Concave-Convex Procedure (CCCP)

---

- 1: Set  $t = 0$  and initialize  $\mathbf{w}_0$
  - 2: **repeat**
  - 3:   Find hyperplane  $\mathbf{v}_t$  such that  $-g(\mathbf{w}) \leq -g(\mathbf{w}_t) + (\mathbf{w} - \mathbf{w}_t) \cdot \mathbf{v}_t$  for all  $\mathbf{w}$
  - 4:   Solve  $\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \mathbf{w} \cdot \mathbf{v}_t$
  - 5:   Set  $t = t + 1$
  - 6: **until**  $[f(\mathbf{w}_t) - g(\mathbf{w}_t)] - [f(\mathbf{w}_{t-1}) - g(\mathbf{w}_{t-1})] < \epsilon$
- 



# Problème d'optimisation du Latent Structural SVM

---

## Algorithm 1 Concave-Convex Procedure (CCCP)

---

- 1: Set  $t = 0$  and initialize  $\mathbf{w}_0$
  - 2: **repeat**
  - 3:   Find hyperplane  $\mathbf{v}_t$  such that  $-g(\mathbf{w}) \leq -g(\mathbf{w}_t) + (\mathbf{w} - \mathbf{w}_t) \cdot \mathbf{v}_t$  for all  $\mathbf{w}$
  - 4:   Solve  $\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \mathbf{w} \cdot \mathbf{v}_t$
  - 5:   Set  $t = t + 1$
  - 6: **until**  $[f(\mathbf{w}_t) - g(\mathbf{w}_t)] - [f(\mathbf{w}_{t-1}) - g(\mathbf{w}_{t-1})] < \epsilon$
- 

4 : résoud le problème d'optimisation standard du Structural SVM

# Problème d'optimisation du Latent Structural SVM

---

**Algorithm 1** Concave-Convex Procedure (CCCP)
 

---

- 1: Set  $t = 0$  and initialize  $\mathbf{w}_0$
  - 2: **repeat**
  - 3:   Find hyperplane  $\mathbf{v}_t$  such that  $-g(\mathbf{w}) \leq -g(\mathbf{w}_t) + (\mathbf{w} - \mathbf{w}_t) \cdot \mathbf{v}_t$  for all  $\mathbf{w}$
  - 4:   Solve  $\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \mathbf{w} \cdot \mathbf{v}_t$
  - 5:   Set  $t = t + 1$
  - 6: **until**  $[f(\mathbf{w}_t) - g(\mathbf{w}_t)] - [f(\mathbf{w}_{t-1}) - g(\mathbf{w}_{t-1})] < \epsilon$
- 

$$\begin{aligned} \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^n & \left( \max_{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y} \times \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}, \mathbf{h}) \rangle + \Delta(\mathbf{y}_i, \hat{\mathbf{y}}(\mathbf{x}_i; \mathbf{w}), \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{w})) \right) \\ & - \frac{C}{n} \sum_{i=1}^n \left( \max_{\mathbf{h} \in \mathcal{H}} \langle \mathbf{w}, \Psi(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}^*(\mathbf{x}_i; \mathbf{w})) \rangle \right) \end{aligned}$$

# Plan

1 Apprentissage Interactif pour la RI

2 Apprentissage faiblement supervisé

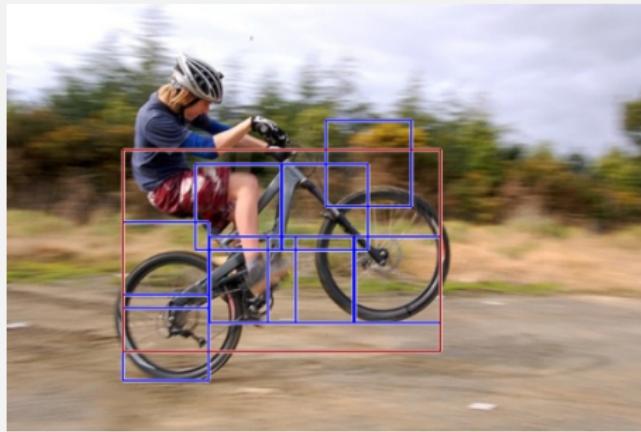
- Contexte
- Latent Structural SVM (LSSVM)
- Exemples d'applications en vision

3 Apprentissage avec des données "exotiques"

# LSSVM : applications en vision

## Deformable Part Model (DPM) [FGMR10]

- DPM : appliqué à la détection d'objets :
  - Problème de classification binaire  $\mathcal{Y} = \pm 1$
  - En entrée : une Bounding Box de l'objet
  - Variable latente : position de (sous)-parties de l'objet



# LSSVM : applications en vision

## Localisation faiblement supervisée [KPK10, RLYFF12, BNFG13]

- $x$  : image
- $h$ , position de la région
- $y$  : classe, *i.e.* "jumping", etc
- Feature map : cf classification multi-classes

### Action Classification

Input  $x$

Annotation  $y$

Latent  $h$



# LSSVM : applications en vision

## Localisation faiblement supervisée

- $x$  : image
- $h$ , position de la région
- $y$  : image pertinent/non pertinente, i.e.  $\mathcal{Y} = \pm 1$  (pour "jumping", etc)
- Feature map : cf ranking

### Action Classification

Input  $x$

Annotation  $y$

Latent  $h$

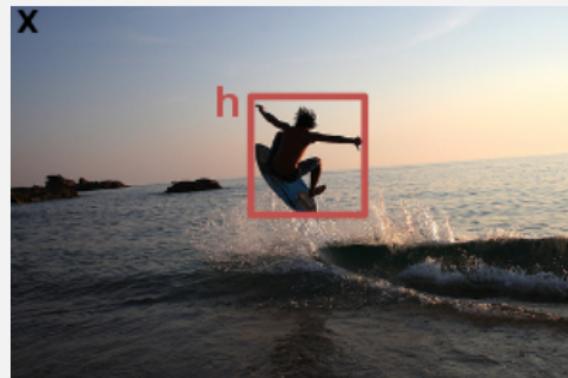
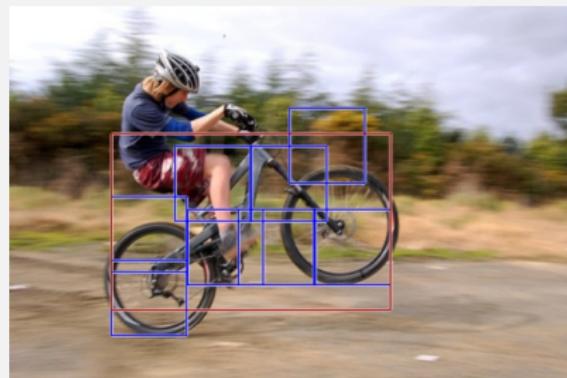


$y = \text{"jumping"}$

# LSSVM : applications en vision

## Localisation faiblement supervisée : bilan

- DPM état de l'art pour les tâches de détection
- Mais l'application du LSSVM pour de la localisation faiblement supervisé dans des tâches de classification de scènes au eu moins de succès
  - Dépendence des parties au tout beaucoup moins rigide



# Localisation faiblement supervisée : extensions récentes

## Extensions du LSSVM

- Optimisation : problème non convexe  $\Rightarrow$  difficulté pour atteindre un minimum local pertinent :
  - Idée du Curriculum learning : apprendre d'abord les paramètres du modèles avec des exemples faciles.  
Challenge : définition d'un exemple facile.
  - Variante : exploration incrementale de l'espace latent : [RLYFF12, BNVG13]
- Modèle : Comment agréger les scores des variables latentes
  - LSSVM : max, i.e. sélection "dure" de la meilleure variable latente
  - HCRF sum [QWM<sup>+</sup>07]
  - Modéliser l'ambiguité entre variables latentes dans modèle LSSVM : M3E [MKP<sup>+</sup>12, MKP<sup>+</sup>12]
  - Modèle unifiés entre HCFR (sum) et LSSVM (max) :  $\epsilon$ -extension [SHPU12], marginal SVM [PLI14]

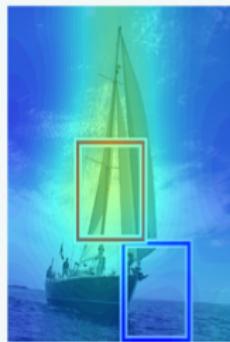
# Localisation faiblement supervisée : extensions récentes

## Extensions du LSSVM : MANTRA [DTC15]

- MANTRA : Minimum mAximum lateNt sTRucturAI SVM
- Utiliser le score  $\max + \min$ , intuition du  $\min$  :
  - Images positives : interdire un score trop faible  $\forall$  région
  - Images négatives : chercher des indicateurs de la non-classe
- Inference et loss-augmented inference exactes pour du ranking



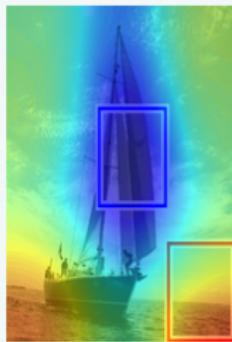
sailing



badminton



bocce



rowing

# LSSVM : applications en vision

## Segmentation faiblement supervisée [XSU14]

- $x$  : image
- Feature map  $\Psi$   
et fonction de  
coût  $\Delta$  : voir  
segmentation



- $y$  : label present in the image :  $\mathbf{y} = \{y_1, \dots, y_i, \dots, y_C\}$ , avec  $y_i \in \{0; 1\}$
- $h$  : label de chaque superpixel :  $h_j \in \{1; C\}$

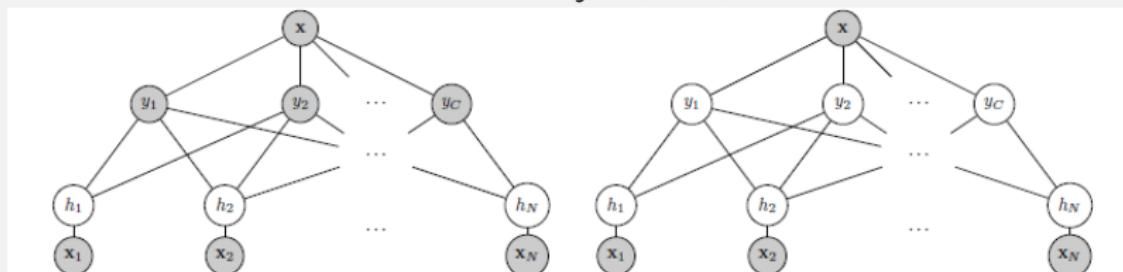


Figure 2. Graphical Model: (Left) Graphical model for learning as well as inference when the tags are provided at test time. (Right) Graphical model for inference when the tags are not provided at test time.

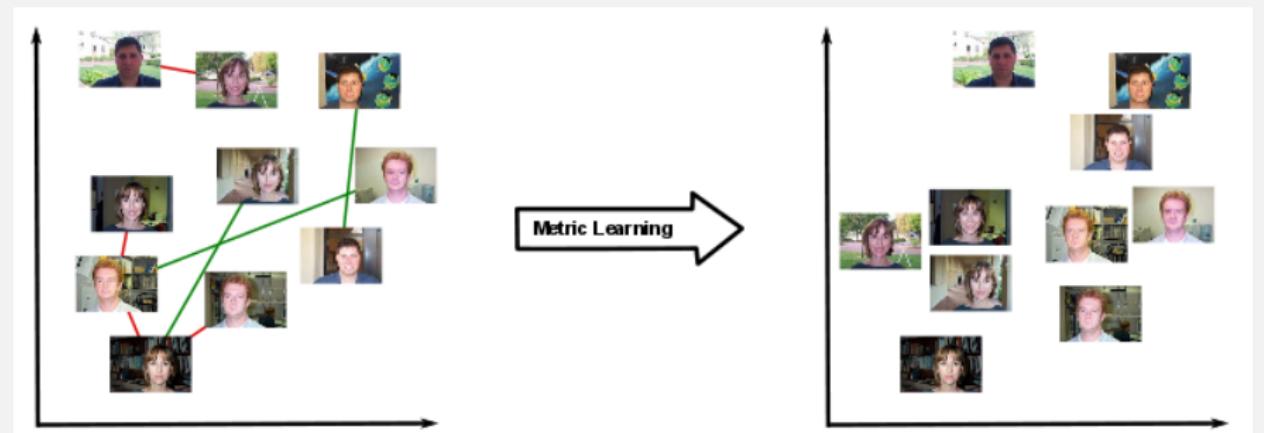
# Plan

- 1 Apprentissage Interactif pour la RI
- 2 Apprentissage faiblement supervisé
- 3 Apprentissage avec des données "exotiques"
  - Attributs
  - Information privilégiée

# Apprentissage à partir d'Attributs

## Motivation

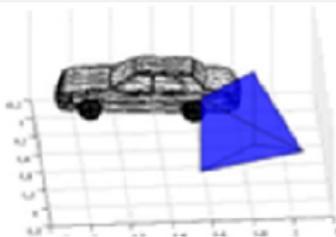
- Données images par toujours uniquement annotées avec information de classe
  - Ex : labels similaire/dissimilaire
  - Voir prochain cours sur l'apprentissage de distance



# Apprentissage à partir d'Attributs

## Motivation

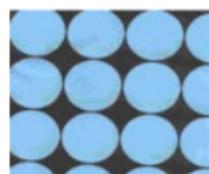
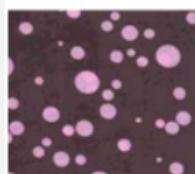
- Données images par toujours uniquement annotées avec information de classe
  - Annotation avec différents niveau de granularité : label, bounding box, segmentation, données 3D (modèles 3D kineck), données eye-tracker, données RGB-D, etc
  - Différentes images peuvent être annotées avec différents types de description



# Apprentissage à partir d'Attributs

## Motivation

- Données images par toujours uniquement annotées avec information de classe
  - Annotation à partir d'attributs (e.g. tags) : red, stripes, round, etc



# Plan

- 1 Apprentissage Interactif pour la RI
- 2 Apprentissage faiblement supervisé
- 3 Apprentissage avec des données "exotiques"
  - Attributs
  - Information privilégiée

# Apprentissage à partir d'Attributs

## Attributs binaires

- Attributs binaires : concepts mid-level
  - Sémantique > descripteurs mais < classes
  - Attributs peuvent être partagés entre classes
- présence / absence d'une propriété dans l'image
- Apprendre et prédiction présence d'attributs :
  - ⊕ Mieux généraliser
  - ⊕ Zero-shot learning

### otter

black: yes  
white: no  
brown: yes  
stripes: no  
water: yes  
eats fish: yes



### polar bear

black: no  
white: yes  
brown: no  
stripes: no  
water: yes  
eats fish: yes



### zebra

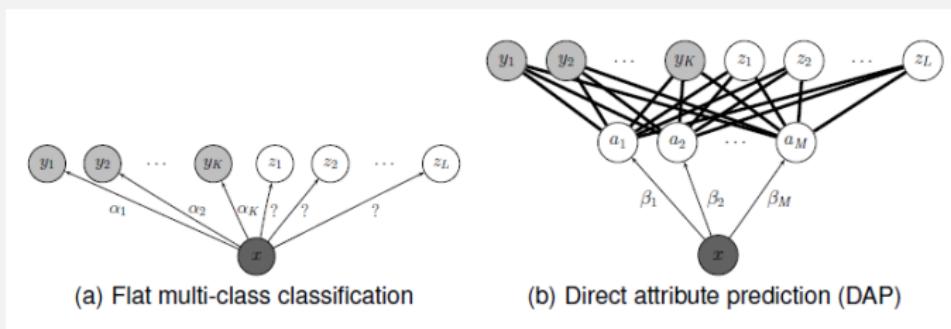
black: yes  
white: yes  
brown: no  
stripes: yes  
water: no  
eats fish: no



# Apprentissage à partir d'Attributs

## Zero-shot learning [LNH09]

- Entrée : exemples  $x$  annotées pour certaines classes  $y_1, \dots, y_N$  et relation entre attributs et toutes les classes  $y_1, \dots, y_N$  et  $z_1, \dots, z_L$
- Apprentissage d'attributs  $a_1, \dots, a_M$  sur une base de données labellée  $\Rightarrow \beta_m$
- Sur de nouvelles données en test pour lesquelles le label n'a pas été vu ( $\in \{z_i\}$ ),  $\Rightarrow$  vecteur d'attributs  $\Rightarrow$  classe



# Apprentissage à partir d'Attributs

## Attributs relatifs [PG11]

- Annotation binaire parfois délicate
- Plus facile de classer deux images / niveau présence attribut
- Apprentissage : pour chaque attribut, apprendre un vecteur  $w$  pour satisfaire les relations  $\sim$  ou  $<$
- Relation avec apprentissage de distance (prochain cours)

### Binary Attributes



Young: Yes  
Smiling: No



Young: Yes  
Smiling: Yes



Young: Yes  
Smiling: Yes



Young: No  
Smiling: Yes



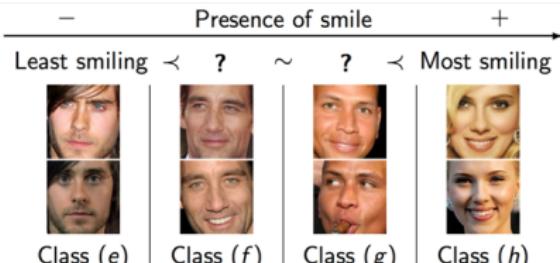
Young: Yes  
Smiling: No

### Relative Attributes

Young



Smiling



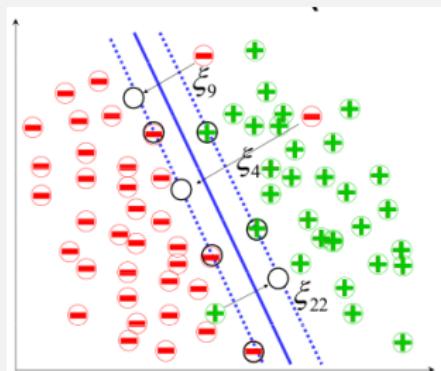
# Plan

- 1 Apprentissage Interactif pour la RI
- 2 Apprentissage faiblement supervisé
- 3 Apprentissage avec des données "exotiques"
  - Attributs
  - Information privilégiée

# Rappel : SVM

## Formulation

- On suppose d'un ensemble de  $N$  exemples d'apprentissage  $\{(x_i, y_i)\}_{i=1,\dots,N}$ , avec  $(x_i, y_i) \in \mathbb{R}^d \times \{-1; 1\}$
- Trouver un hyperplan  $w$  minimisant la fonction objective suivante :



$$\begin{aligned} & \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ & \text{s.t. } \forall i = \{1, \dots, N\} \\ & \quad y_i [ \langle w, x_i \rangle + b ] \geq 1 - \xi_i \\ & \quad \xi_i \geq 0 \end{aligned} \tag{3}$$

# Information privilégiée

## SVM+ [VV09, PV10]

- On suppose disposer d'information supplémentaire **uniquement pendant l'apprentissage**  $\Rightarrow x_i^* = \text{information privilégiée}$
- Problème d'optimisation SVM+ [VV09, PV10] :

$$\begin{aligned} & \min_{w \in \mathbb{R}^d, b \in \mathbb{R}, w^* \in \mathbb{R}^{d^*}, b^* \in \mathbb{R}} \frac{1}{2} (\|w\|^2 + \gamma \|w^*\|^2) + C \sum_{i=1}^N \langle w^*, x_i^* \rangle + b^* \\ & \text{s.t. } \forall i = \{1, \dots, N\} \\ & \quad y_i [\langle w, x_i \rangle + b] \geq 1 - [\langle w^*, x_i^* \rangle + b^*] \\ & \quad \langle w^*, x_i^* \rangle + b^* \geq 0 \end{aligned} \tag{4}$$

- Objectif : améliorer la qualité du modèle  $w$  appris,
  - Meilleur performance de généralisation du  $w$  appris avec l'information privilégiée, sur des données de test où cette info n'est pas disponible

# Information privilégiée

## SVM+ : lien avec SVM

SVM

$$\begin{aligned} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ y_i [\langle w, x_i \rangle + b] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{aligned} \quad (5)$$

SVM+

$$\begin{aligned} \frac{1}{2} (\|w\|^2 + \gamma \|w^*\|^2) + C \sum_{i=1}^N \langle w^*, x_i^* \rangle + b^* \\ y_i [\langle w, x_i \rangle + b] \geq 1 - [\langle w^*, x_i^* \rangle + b^*] \\ \langle w^*, x_i^* \rangle + b^* \geq 0 \end{aligned} \quad (6)$$

- Soft-margin SVM : optimisation des  $\xi_i$  augmente le nombre de paramètres / hard-margin  $\Rightarrow$  plus d'exemples nécessaires pour converger [VV09] :
  - Converge en  $1/\sqrt{N}$  en soft-margin vs  $1/N$  en hard margin
  - MAIS si on suppose qu'un oracle connaît la valeur des  $\xi_i$  optimaux  $\Rightarrow$  convergence réduite à  $1/N$
- Idée du SVM+ : l'information privilégiée fournit un substitut (proxy) à l'oracle pour estimer les  $\xi_i$  optimaux

# Information privilégiée

## SVM+ : utilisation en vision [SQL13]

$x$  : image



$x^*$  : attributes

black:	no
white:	yes
brown:	no
stripes:	no
water:	yes
eats fish:	yes

$x$  : image



$x^*$  : rationale



$x$  : image



$x^*$  : bounding box



$x$  : image



$x^*$  : text

Apples and Honey  
With Pomegranates  
Rosh Hashanna  
(Jewish New Year)  
Meal

# References I

-  H. Bilen, V.P. Namboodiri, and L.J. Van Gool, *Object classification with latent window parameters*, International Journal of Computer Vision, 2013.
-  Klaus Brinker, *Incorporating diversity in active learning with support vector machines*, In Proceedings of the 20th International Conference on Machine Learning, AAAI Press, 2003, pp. 59–66.
-  Thibaut Durand, Nicolas Thome, and Matthieu Cord, *MANTRA : Minimum Maximum Latent Structural SVM for Image Classification and Ranking*, ICCV, 2015.
-  P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, *Object detection with discriminatively trained part based models*, IEEE Transactions on Pattern Analysis and Machine Intelligence **32** (2010), no. 9, 1627–1645.
-  P. Kumar, B. Packer, and D. Koller, *Self-paced learning for latent variable models*, Advances in Neural Information Processing Systems (NIPS 2010), 2010.
-  Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling, *Learning to detect unseen object classes by betweenclass attribute transfer*, In CVPR, 2009.
-  Kevin Miller, M. Pawan Kumar, Benjamin Packer, Danny Goodman, and Daphne Koller, *Max-margin min-entropy models*, Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS, 2012.
-  D. Parikh and K. Grauman, *Relative attributes*, ICCV, 2011.
-  Wei Ping, Qiang Liu, and Alex Ihler, *Marginal structured svm with hidden variables*, Proceedings of the 31st International Conference on Machine Learning (ICML-14) (Tony Jebara and Eric P. Xing, eds.), JMLR Workshop and Conference Proceedings, 2014, pp. 190–198.

# References II

-  Dmitry Pechony and Vladimir Vapnik, *On the theory of learnining with privileged information*, Advances in Neural Information Processing Systems 23 (J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, eds.), Curran Associates, Inc., 2010, pp. 1894–1902.
-  Ariadna Quattoni, Sybor Wang, Louis-Philippe Morency, Michael Collins, and Trevor Darrell, *Hidden conditional random fields*, IEEE Trans. Pattern Anal. Mach. Intell. **29** (2007), no. 10, 1848–1852.
-  Olga Russakovsky, Yuanqing Lin, Kai Yu, and Li Fei-Fei, *Object-centric spatial pooling for image classification*, ECCV, 2012.
-  Alexander G. Schwing, Tamir Hazan, Marc Pollefeys, and Raquel Urtasun, *Efficient structured prediction with latent variables for general graphical models*, Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012, 2012.
-  V. Sharmanska, N. Quadrianto, and C.H. Lampert, *Learning to rank using privileged information*, International Conference on Computer Vision (ICCV), 2013.
-  Simon Tong and Daphne Koller, *Support vector machine active learning with applications to text classification*, J. Mach. Learn. Res. **2** (2002), 45–66.
-  Vladimir Vapnik and Akshay Vashist, *A new learning paradigm : Learning using privileged information.*, Neural Networks **22** (2009), no. 5-6, 544–557.
-  Jia Xu, Alexander G. Schwing, and Raquel Urtasun, *Tell me what you see and i will show you where it is*, CVPR, 2014.
-  Chun-Nam Yu and T. Joachims, *Learning structural svms with latent variables*, International Conference on Machine Learning (ICML), 2009.

# References III



Alan L. Yuille and Anand Rangarajan, *The concave-convex procedure*, Neural Computation (2003).