Statistical syntax analysis
for signal processing

Olivier Catoni
Olivier.Catoni@ensae.fr
http://ocatoni.perso.math.cnrs.fr/

CREST, CNRS — UMR 9194
Université Paris–Saclay

Conférence en l'honneur de Robert Azencott

Éléments de mathématiques pour l'Intelligence Artificielle

ENS Paris-Saclay, Cachan, May 14, 2019

Joint work with
Gautier Appert

L'inconscient est structuré comme un langage.

Jacques Lacan

General motivation

As the human brain translates its perception of the world into words, we may conjecture that it uses the same kind of information analysis when dealing with signals as when dealing with language. We will explore more specifically the case of visual perception and propose a mathematical framework for the syntax analysis of digital images.

The two principles of syntax analysis

Syntax analysis combines two actions:

1. Grouping
2. Context analysis

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

Image data base

$\Longrightarrow$

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

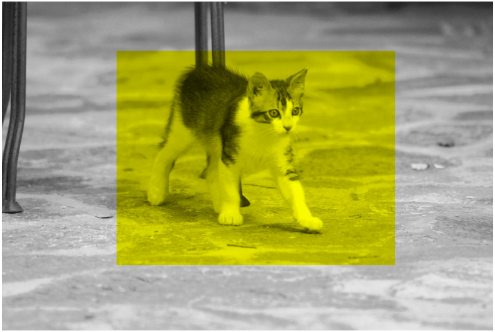


Image data base

$\Longrightarrow$

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

Image data base

$\Longrightarrow$

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

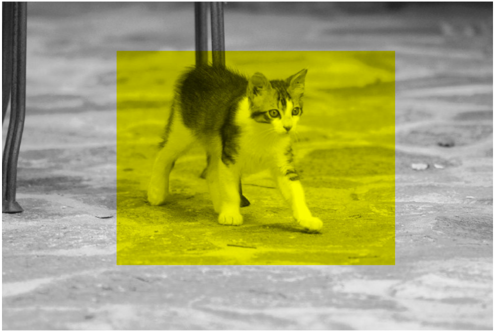


Image data base

$\Longrightarrow$

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

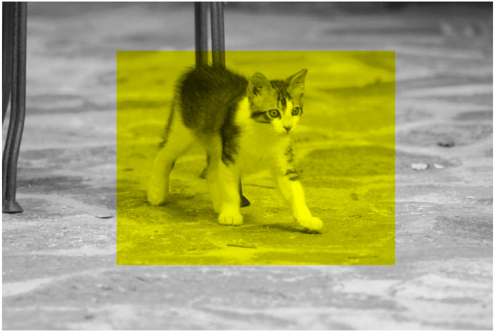


Image data base

$\Longrightarrow$

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

Image data base

⟹

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Input

A training set of digital images of the same size, obtained by extracting randomly located windows from a data base of larger images. The training set is a model for data acquired by some sensor (some retina) looking at the image data base.

Image data base

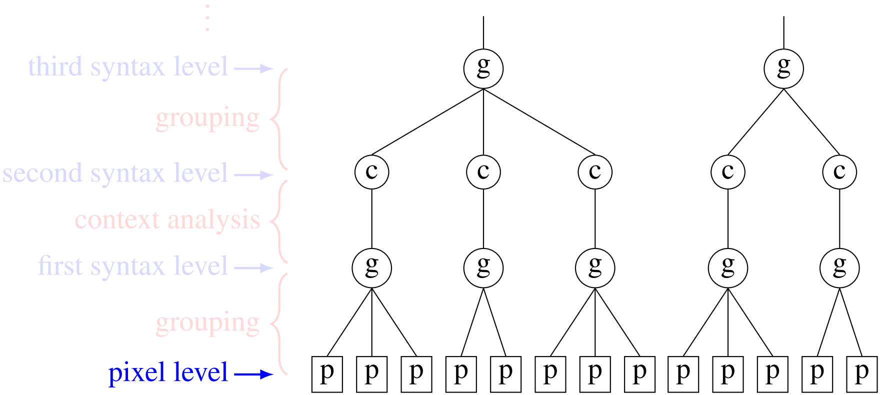
$\Longrightarrow$

Randomly located sensor measurements

A statistical framework for syntax analysis of digital images

Output

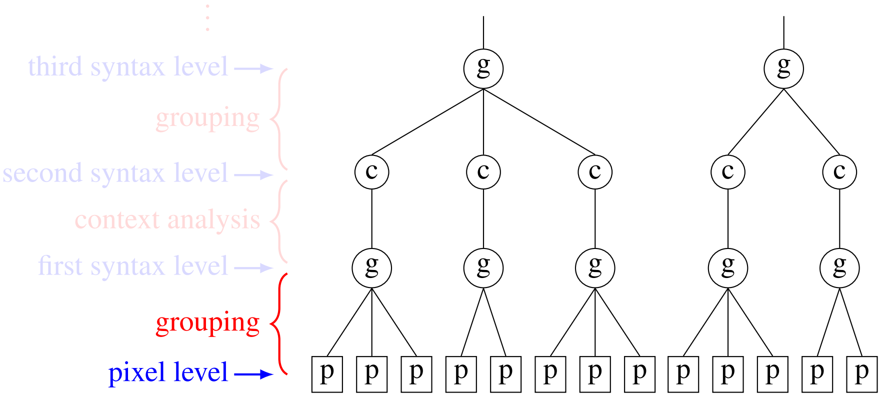
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

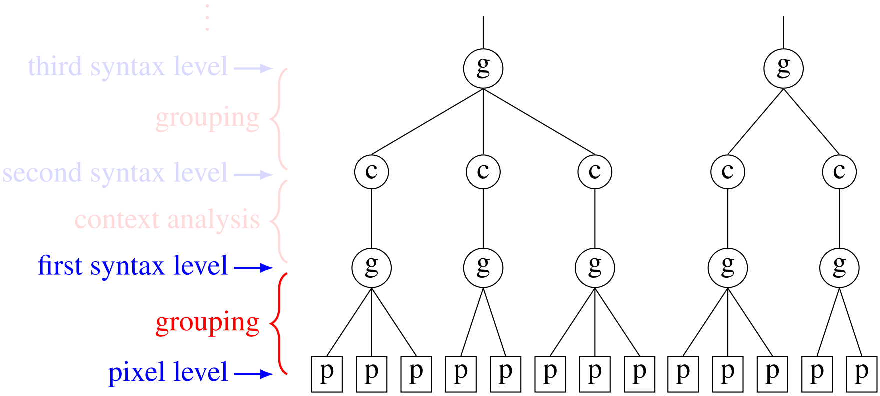
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

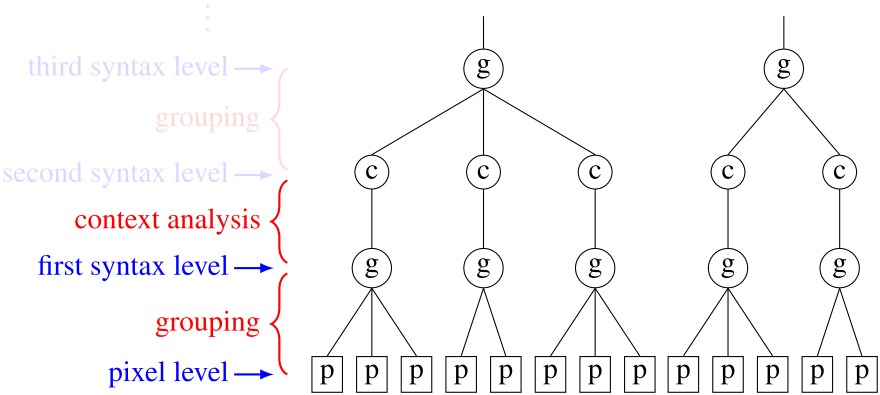
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

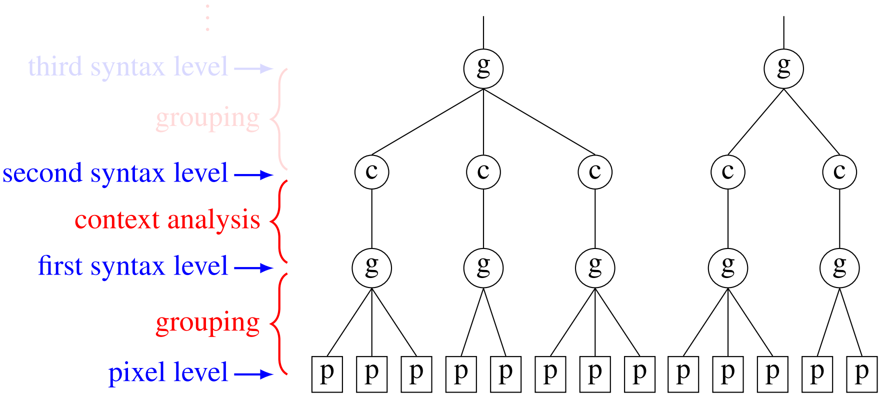
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

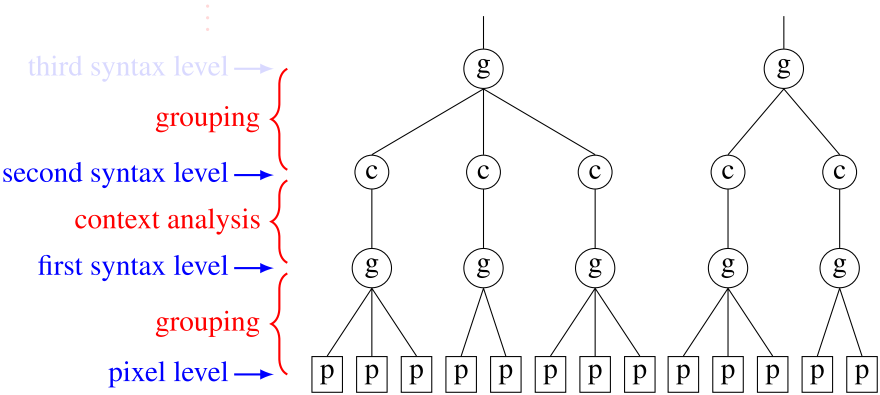
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

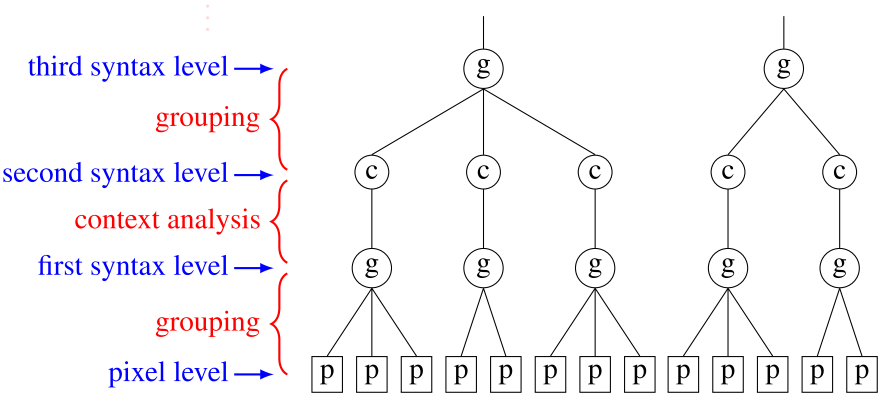
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

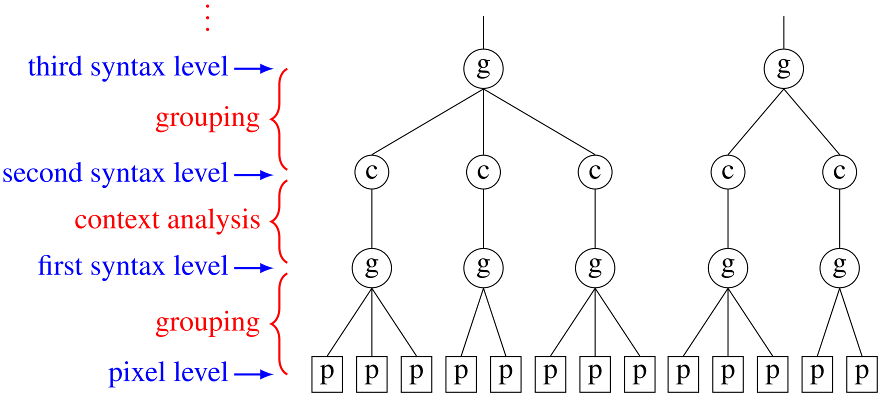
A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Output

A syntax tree for each image of the training set. That is a multiscale unsupervised pixel classification corresponding to nested pixel partitions.



A statistical framework for syntax analysis of digital images

Grouping principle

Cluster together pixels or syntax labels forming a frequent pattern (or in other words appearing in the same images).
Grouping expresses the fact that I can see a_1 and a_2 and $\cdots a_m$ together.

Context analysis principle

Cluster together syntax labels appearing in the same context in different images.
Context analysis expresses the fact that I can see a_1 or a_2 or $\cdots a_m$ in the same contexts.

Syntax trees and logic

Syntax trees express logical formulas of the type

$$\bigwedge_{i_{2m-1}} \cdots \bigwedge_{i_3} \bigvee_{i_2} \bigwedge_{i_1} a_{i_1,\ldots,i_{2m-1}}.$$

We can also see the syntax trees as defining sets of rewriting rules. Thus if we can learn syntax trees, based on some kind of statistical inference, we can learn some kind of unsupervised logical reasoning about the training set.

Mathematical model: pixel grouping

Use probability distributions to represent data

Let $X \in \mathcal{X} \subset \mathbb{R}^d$ be a random image of size d.
Let the training set be made of n independent copies of X.
Represent the content of image X by the conditional probability distribution

$$\mathbb{P}_{S,V \mid X} = \frac{1}{d} \sum_{s=1}^{d} \delta_s \otimes \mathcal{N}(X_s, \sigma^2),$$

where $S \in [\![1, d]\!]$ and $V \in \mathbb{R}$ are two additional random variables and the above defines the joint distribution of the triplet (X, S, V). The image X is a function of its new representation $\mathbb{P}_{S,V \mid X}$ since

$$X_s = \int v \, \mathrm{d}\mathbb{P}_{V \mid X,\, S=s}(v).$$

The variance parameter σ is a free parameter of the new representation of X.

Mathematical model: pixel grouping

Grouping and conditional independence

We want a pixel classification function $\ell : \mathcal{X} \times [\![1,d]\!] \longrightarrow [\![1,k]\!]$ such that

$$\mathbb{P}_{X,\, S,\, V \,|\, \ell(X,S)} \simeq \mathbb{P}_{X \,|\, \ell(X,\, S)} \otimes \mathbb{P}_{S,\, V \,|\, \ell(X,\, S)}.$$

When this is true each image is partitioned into pixel patterns:

$$\mathbb{P}_{S,\, V \,|\, X} \simeq \sum_{j=1}^{k} \mathbb{P}\big(\ell(X,S) = j \,|\, X\big) \; \underbrace{\mathbb{P}_{S,\, V \,|\, \ell(X,\, S) = j}}_{\text{pattern number } j}.$$

Mathematical model: pixel grouping

Using a model and a loss function

Rather than checking the desired property of the previous slide, we introduce a model and a loss function.

$$\mathcal{C}(\ell) = \inf_{Q \in \mathcal{Q}_\ell} \mathcal{K}(Q_{X,S,V}, \mathbb{P}_{X,S,V}).$$

Motivation: according to Sanov's theorem

$$- \inf_{Q \in \mathring{\Gamma}} \mathcal{K}(Q, \mathbb{P}) \leq \liminf_{n \to \infty} \frac{1}{n} \log \Big[\mathbb{P}^{\otimes n} (\overline{\mathbb{P}}_n \in \Gamma) \Big]$$
$$\leq \limsup_{n \to \infty} \frac{1}{n} \log \Big[\mathbb{P}^{\otimes n} (\overline{\mathbb{P}}_n \in \Gamma) \Big] \leq - \inf_{Q \in \overline{\Gamma}} \mathcal{K}(Q, \mathbb{P}),$$

meaning that $\mathcal{K}(Q, \mathbb{P})$ reflects the log likelihood of observing an empirical measure close to Q. In other words we want to maximize the likelihood of the model under the data distribution. In particular we want $Q \ll \mathbb{P}$.

Mathematical model: pixel grouping

More motivation

The expression

$$\mathcal{K}(Q, \mathbb{P}) = \int \left[1 - \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}} + \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}} \log\left(\frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}} \right) \right] \mathrm{d}\mathbb{P}$$

also shows that the criterion penalizes more the fact that $\frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}$ is large than the fact that it is small: we do not want events that are likely under the model and unlikely under the data distribution, and to achieve this, we are ready to accept that some events may be less likely under the model distribution than under the data distribution.

Mathematical model: pixel grouping

Choose a model that satisfies the conditional independence assumption

Consider

$$\mathcal{Q}_\ell = \Big\{ Q \,:\, Q_{S,\,V\,|\,X,\ell(X,\,S)} = Q_{S,\,V\,|\,\ell(X,\,S)} \Big\}.$$

Try to compute the loss function

Introduce the random variable $W = \ell(X,S)$.

$$\mathcal{C}(\ell) \overset{\text{def}}{=} \inf_{Q \in \mathcal{Q}_\ell} \mathcal{K}(Q_{X,\,S,\,V}, \mathbb{P}_{X,\,S,\,V})$$

$$= -\log\left\{ \sup_{Q_{X\,|\,W}} \mathbb{P}_{W,\,S}\left[\exp\left(-\mathcal{K}(Q_{X\,|\,W}, \mathbb{P}_{X\,|\,W,\,S}) - \frac{1}{2\sigma^2}\,\mathbf{Var}\big(Q_{X_S\,|\,W}\big)\right)\right]\right\}$$

Mathematical model: pixel grouping

Make life simpler by making the model smaller

Considering $A_j = \operatorname{supp}(Q_{X \mid W=j})$ and $B_j = \operatorname{supp}(Q_{S \mid W=j})$, we see that $A_j \times B_j = \operatorname{supp}(Q_{X,\, S \mid W=j})$ when $Q \in \mathcal{Q}_\ell$ and that Q remains in $\mathcal{Q}_\ell$ if we change ℓ so that $\ell^{-1}(j) = A_j \times B_j$. We will accordingly assume without loss of generality that $\ell^{-1}(j) = A_j \times B_j$ for some family $A_j \times B_j$ of disjoint product sets. Since $\sup_{Q_{X \mid W}}$ is not easy to compute in the previous formula, we will also reduce the model to

$$\mathcal{Q}_\ell = \Big\{ Q \,:\, Q_{S,\, V \mid X,\, W} = Q_{S,\, V \mid W} \text{ and } Q_{X \mid W} = \mathbb{P}_{X \mid W} \Big\},$$

where $W = \ell(X, S)$. With these modifications $\mathcal{K}(Q_{X \mid W}, \mathbb{P}_{X \mid W,\, S}) = 0$.

Mathematical model: pixel grouping

Decomposing the criterion into block weights

We can write $\mathcal{C}(\ell) = -\log\left(\sum_{j=1}^{k} \mathcal{W}(\ell,j)\right)$, where

$$\begin{aligned}\mathcal{W}(\ell,j) &= \mathbb{P}(X \in A_j)\mathbb{P}(S \in B_j)\mathbb{P}_{S \mid S \in B_j}\left[\exp\left(-\frac{1}{2\sigma^2}\,\mathbf{Var}\left(\mathbb{P}_{X_S \mid X \in A_j}\right)\right)\right] \\ &= \mathcal{W}(A_j,B_j).\end{aligned}$$

Thus $\mathcal{W}(\ell,j)$ depends only on A_j, an image subset, and B_j, a pixel subset, or pixel pattern. So we are led to look for disjoint product sets $A_j \times B_j$, $1 \leq j \leq k$, maximizing

$$\sum_{j=1}^{k} \mathcal{W}(A_j,B_j).$$

Notice that we arrive at a criterion that reminds of the k-means criterion.

Mathematical model: pixel grouping

From large and unique towards small and frequent patterns

Assume for simplicity that $\mathcal{X} = \{x_1, \dots, x_N\}$ is finite. Instead of maximizing $\sum_{j=1}^{k} \mathcal{W}(A_j, B_j)$ for k fixed we will start from the trivial solution

$$k = N, \quad A_j = \{x_j\}, \quad B_j = \{1, \dots, d\}$$

that satisfies $\sum_{j=1}^{k} \mathcal{W}(A_j, B_j) = 1$ and create from there by induction a sequence of solutions $\mathcal{S}_k = \{A_{k,j}, B_{k,j} \,:\, 1 \le j \le k\}$ for $k = N, N+1, N+2, \dots,$ trying at each step to maximize the new weight $\mathcal{W}(A_{k,k}, B_{k,k})$ while keeping $\mathcal{W}(\mathcal{S}_k) = \sum_{j=1}^{k} \mathcal{W}(A_{k,j}, B_{k,j})$ above a certain level.

Mathematical model: pixel grouping

Induction step: pattern splitting and jigsaw puzzles

Starting from $\mathcal{S}_k$, we choose a pair of indices $J \subset [\![1,k]\!]$ and define $\mathcal{S}_{k+1}$ by

$$
\begin{aligned}
A_{k+1,k+1} &= \bigcup_{j \in J} A_{k,j}, \\
B_{k+1,k+1} &= \left\{ s \in \bigcap_{j \in J} B_{k,j} \;:\; \mathbf{Var}\left(\mathbb{P}_{X_s \mid X \in A_{k+1,k+1}}\right) \leq a \right\}, \\
A_{k+1,j} &= A_{k,j}, \quad 1 \leq j \leq k \\
B_{k+1,j} &= \begin{cases} B_{k,j} \setminus B_{k+1,k+1}, & j \in J, \\ B_{k,j}, & j \in [\![1,k]\!] \setminus J. \end{cases}
\end{aligned}
$$

Doing so we are sure that

$$\mathcal{W}(\mathcal{S}_k) \geq \exp\left(-\frac{a}{2\sigma^2}\right).$$

Mathematical model: pixel grouping

Choice of the pair J

We choose the pair $J = \{i,j\}$ to make $\mathcal{W}(A_{k+1}, B_{k+1})$ as large as possible, while keeping the computation cost reasonable. We propose the following choice

$$i \in \arg \max_{i \in [\![1,\, k]\!]} \mathcal{W}(A_{k,i}, B_{k,i}), \quad j \in \arg \max_{j \in [\![1,\, k]\!] \setminus \{i\}} \mathcal{W}(A_{k+1,k+1}, B_{k+1,k+1}).$$

Mathematical model: label grouping

Loss function and model choice

Consider random images X described by label probability measures $\mathbb{P}_{Y\,|\,X}$, where $Y \in \mathcal{Y}$ a finite set of labels (in the building of the syntax tree, these will be the labels at the previous syntax level). For any label classification function $\ell : \mathcal{X} \times \mathcal{Y} \to [\![1,k]\!]$, consider the loss function

$$\mathcal{C}(\ell) = \inf_{Q \in \mathcal{Q}_\ell} \mathcal{K}(Q_{X,Y}, \mathbb{P}_{X,Y}),$$

where

$$\mathcal{Q}_\ell = \Big\{ Q : Q_{Y\,|\,X,\,\ell(X,Y)=j} = \underbrace{Q_{Y\,|\,\ell(X,\,Y)\,=j}}_{\text{pattern number } j} \Big\}.$$

Putting $W = \ell(X,Y)$, we can prove that

$$\mathcal{C}(\ell) = -\log \sup_{Q_{X\,|\,W}} \mathbb{P}_{Y,\,W}\Big[\exp\Big(-\mathcal{K}(Q_{X\,|\,W}, \mathbb{P}_{X\,|\,Y,\,W})\Big)\Big]$$

Mathematical model: label grouping

From classification functions to product sets

Remarking that, for any $Q \in \mathcal{Q}_\ell$,

$$\operatorname{supp}\left(Q_{X,\,Y\,|\,W=j}\right) = \underbrace{\operatorname{supp}\left(Q_{X\,|\,W=j}\right)}_{=A_j} \times \underbrace{\operatorname{supp}\left(Q_{Y\,|\,W=j}\right)}_{=B_j} \subset \ell^{-1}(j),$$

we can introduce

$$\mathcal{A} = \left\{(A_j, B_j)_{j=1}^k \;:\; (A_i \times B_i) \cap (A_j \times B_j) = \varnothing, i \neq j \in [\![1, k]\!]\right\}$$

and see that

$$\inf_\ell \mathcal{C}(\ell) = -\log \sup_{(A,B)\in\mathcal{A}} \sum_{j=1}^k \mathbb{P}\Big((X,Y) \in A_j \times B_j\Big)$$

$$\times \sup_{Q_{X\,|\,X\in A_j}} \mathbb{P}_{Y\,|\,(X,\,Y)\in A_j\times B_j}\left[\exp\Big(-\mathcal{K}\big(Q_{X\,|\,X\in A_j}, \mathbb{P}_{X\,|\,X\in A_j,\,Y}\big)\Big)\right]$$

$$= -\log \sup_{(A,B)\in\mathcal{A}} \sum_{j=1}^k \mathcal{W}(A_j, B_j).$$

Mathematical model: label grouping

Pattern splitting

Start from the trivial solution

$$\mathcal{S}_N = \Big\{(A_{N,\,j}, B_{N,\,j}),\ 1 \le j \le N\Big\},$$

where $A_{N,\,j} = \{x_j\}$ and $B_{N,j} = \operatorname{supp}\Big(\mathbb{P}_{Y\,|\,X=x_j}\Big)$. Based on $\mathcal{S}_k$, compute $\mathcal{S}_{k+1}$ in the following way. For some pair $J \subset [\![1,k]\!]$, put

$$A_{k+1,\,k+1} = \bigcup_{j\in J} A_{k,\,j}, \qquad A_{k+1,\,j} = A_{k,\,j}$$

$$B_{k+1,\,k+1} = \Big\{y \in \bigcap_{j\in J} B_{k,\,j} \ :\ \mathcal{K}\Big(\mathbb{P}_{X\,|\,(X,\,Y)\in A_{k+1,\,k+1}\times\bigcap_{j\in J} B_{k,\,j}},\ \mathbb{P}_{X\,|\,X\in A_{k+1,\,k+1},\,Y=y}\Big) \le a\Big\},$$

$$B_{k+1,\,j} = \begin{cases} B_{k,\,j} \setminus B_{k+1,\,k+1}, & j \in J, \\ B_{k,\,j}, & j \in [\![1,k]\!] \setminus J. \end{cases}$$

Mathematical model: label grouping

Choice of the pair of patterns

Remark first that whatever the choice of J, $\mathcal{C}(\mathcal{S}_k) \leq a$ for all $k \geq N$. We propose to take $J = \{i, j\}$ where

$$i \in \arg \max_{i \in [\![1,\, k]\!]} \mathcal{W}(A_{k,\, i}, B_{k\, i}), \quad j \in \arg \max_{j \in [\![1,\, k]\!] \setminus \{i\}} \mathcal{W}(A_{k+1,\, k+1}, B_{k+1,\, k+1}),$$

where

$$\mathcal{W}(A, B) = \mathbb{P}_{X,\, Y}(A \times B)$$
$$\mathbb{P}_{Y \,|\, (X,\, Y) \in A \times B}\Big[\exp\Big(-\mathcal{K}(\mathbb{P}_{X \,|\, (X,\, Y) \in A \times B}, \mathbb{P}_{X \,|\, X \in A, Y})\Big)\Big].$$

Mathematical model: context analysis

Defining a context function

Starting from a representation $\mathbb{P}_{Y \mid X}$ of a random image by a label distribution, we merge pairs of labels using the classification function $c(X, Y)$ such that $c^{-1}(j) = A_j \times B_j$, where $|B_j| \in \{1, 2\}$. We require that $\mathbb{P}_{Y \mid X,\, c(X,Y)} \simeq \mathbb{P}_{Y \mid c(X,\, Y)}$. For this we use the criterion

$$\mathcal{C}(c) = \inf_{Q \in \mathcal{Q}_c} \mathcal{K}(Q_{X,\, Y}, \mathbb{P}_{X,\, Y}),$$

where

$$\mathcal{Q}_c = \Big\{ Q \,:\, Q_{Y \mid X,\, c(X,Y)} = Q_{Y \mid c(X,Y)} \Big\}.$$

We use a splitting scheme starting from $B_j = \{y_j\}$ and $A_j = \operatorname{supp}(\mathbb{P}_{X \mid Y = y_j})$, where $\mathcal{Y} = \{y_j,\, 1 \leq j \leq N\}$. We define the context function as

$$f(x, y) = \begin{cases} y, & \text{when } |B_{c(x,y)}| = 1, \\ y' & \text{when } |B_{c(x,y)}| = 2 \text{ and } B_{c(x,y)} = \{y, y'\}. \end{cases}$$

Mathematical model: context analysis

Context analysis

We put $T = f(X,Y)$, to obtain a triplet of random variables (X,Y,T) where T is the context of Y in the random image X. We will now compute a syntax label $g(Y,T)$, based on the distribution of the pair (Y,T). We require that $\mathbb{P}_{Y,T} \simeq \mathbb{P}_{Y \mid g(Y,T)} \otimes \mathbb{P}_{T \mid g(Y,T)}$. To achieve this, we introduce the criterion

$$\mathcal{C}(g) = \inf_{Q \in \mathcal{Q}_g} \mathcal{K}(Q_{Y,T}, \mathbb{P}_{Y,T}),$$

where $\mathcal{Q}_g = \left\{ Q \, : \, Q_{T \mid Y, g(Y,T)} = Q_{T \mid g(Y,T)} \right\}$. We use a splitting scheme starting from $g^{-1}(j) = A_j \times B_j$, where $A_j = \{y_j\}$ and $B_j = \operatorname{supp}(\mathbb{P}_{T \mid Y = y_j})$. The next syntax level is defined as $W = g(Y, f(X,Y))$.

What it looks like

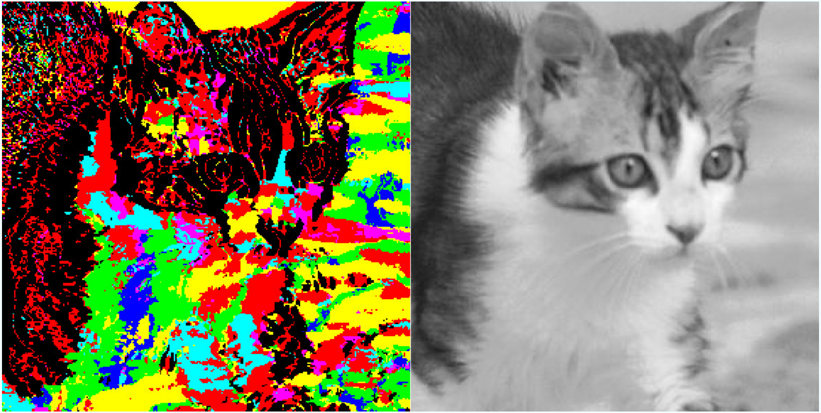
First level patterns may be quite faithfull

What it looks like

First level patterns may be quite faithfull

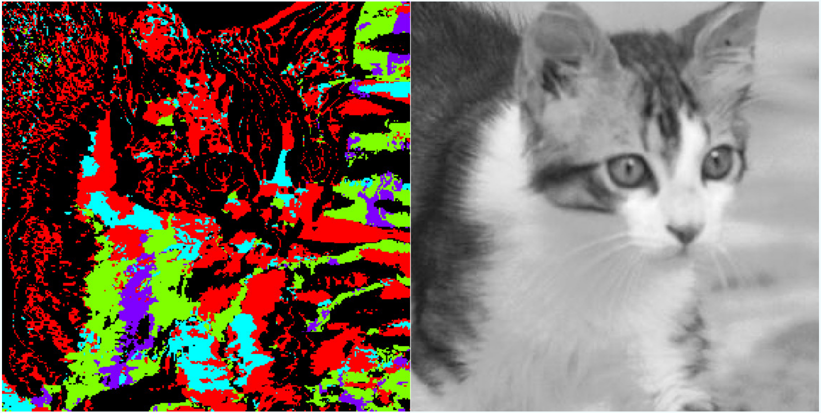
What it looks like

First and higher level label images



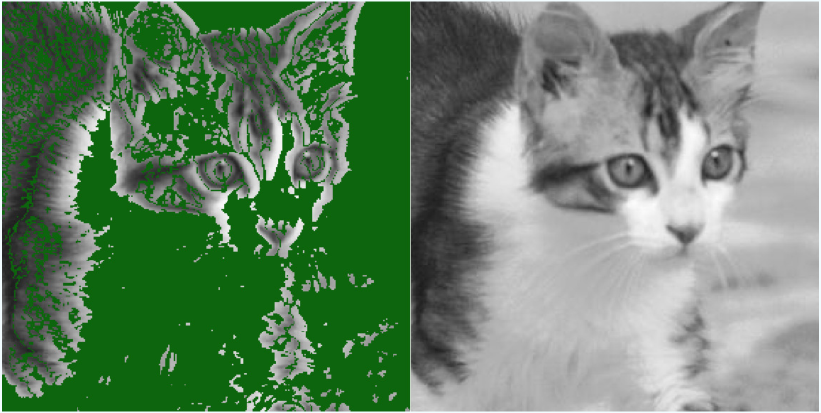
What it looks like

First and higher level label images



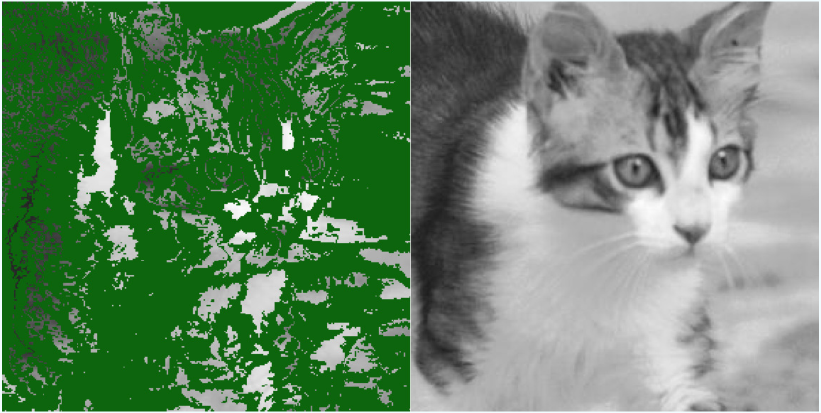
What it looks like

The images sharing the patterns of a given one (first level)



What it looks like

The images sharing the patterns of a given one (first level)

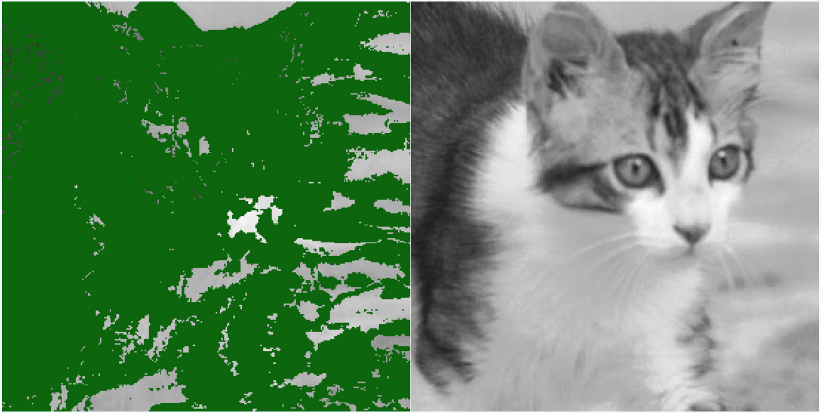


What it looks like

The images sharing the patterns of a given one (first level)

What it looks like

The images sharing the patterns of a given one (first level)



What it looks like

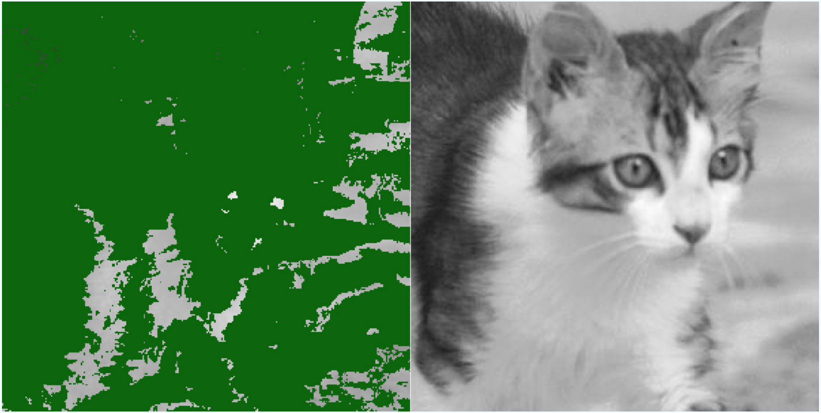
The images sharing the patterns of a given one (first level)

What it looks like

The images sharing the patterns of a given one (first level)

What it looks like

The images sharing the patterns of a given one (first level)

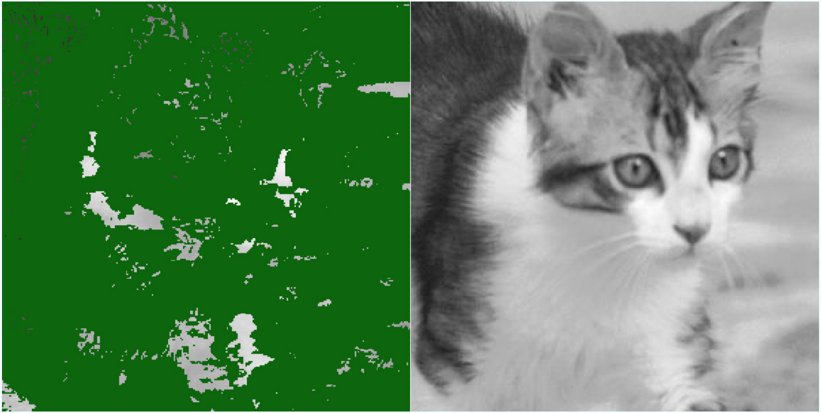


What it looks like

The images sharing the patterns of a given one (first level)

What it looks like

The images sharing the patterns of a given one (first level)



What it looks like

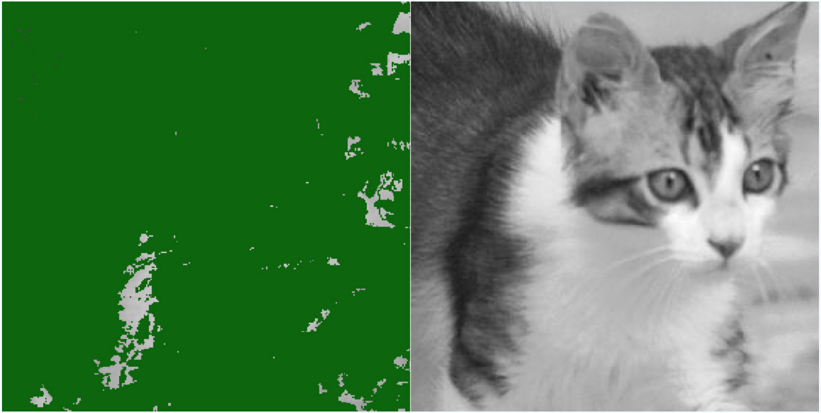
The images sharing the patterns of a given one (first level)

What it looks like

The images sharing the patterns of a given one (first level)

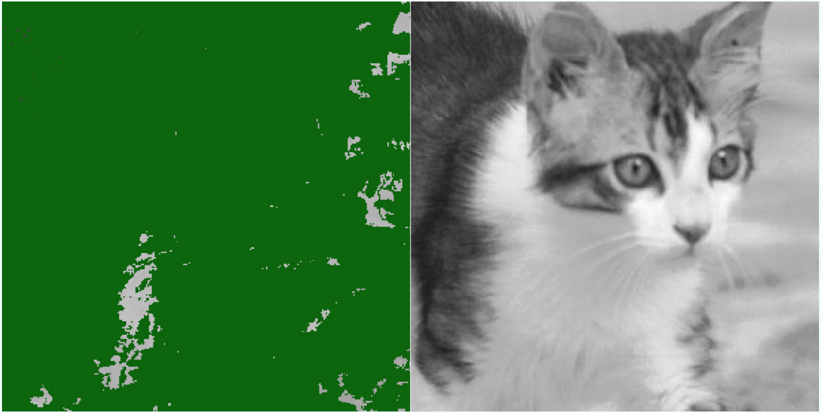
What it looks like

The images sharing the patterns of a given one (first level)



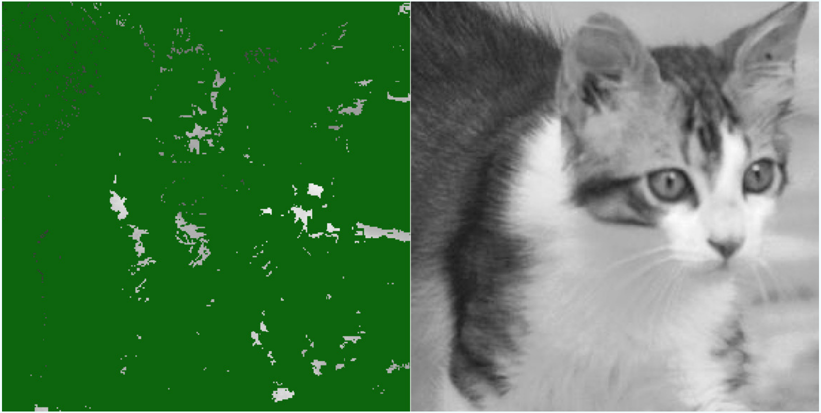
What it looks like

The images sharing the patterns of a given one (first level)



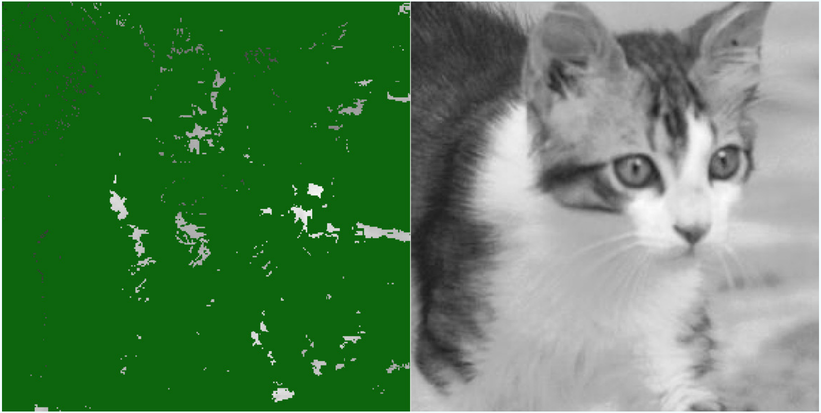
What it looks like

The images sharing the patterns of a given one (first level)



What it looks like

The images sharing the patterns of a given one (first level)

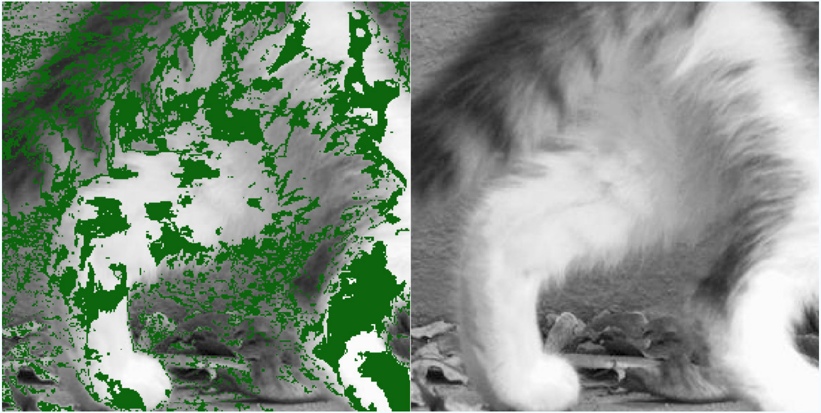


What it looks like

Images sharing some higher level syntax label: a cat's leg label?

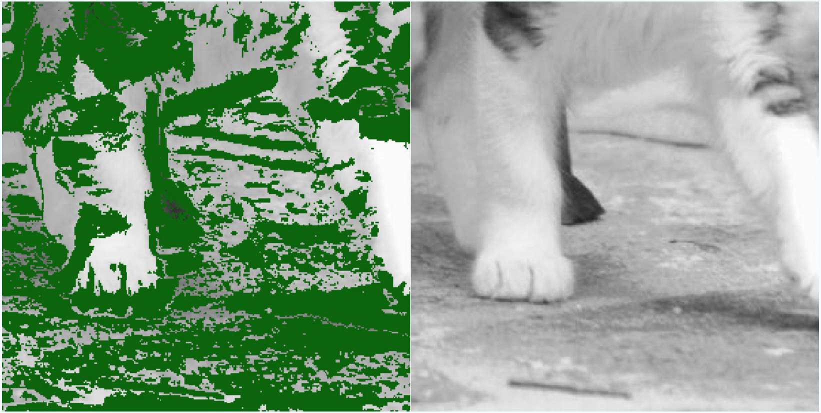
What it looks like

Images sharing some higher level syntax label: a cat's leg label?



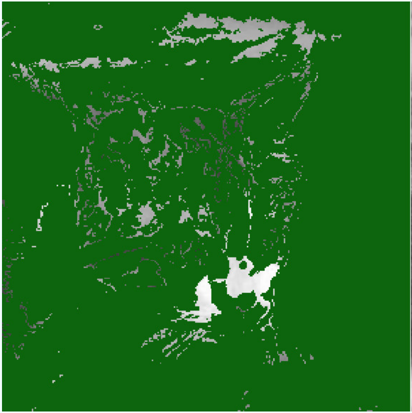
What it looks like

Images sharing some higher level syntax label: a cat’s leg label?



What it looks like

Cat's head matching from a training set made from one hundred random windows from image one and a single window centered on the cat's head in image two.



What it looks like

Cat's head matching from a training set made from one hundred random windows from image one and a single window centered on the cat's head in image two.