Image segmentation: from the Mumford-Shah conjecture to more global variational formulations by deep learning

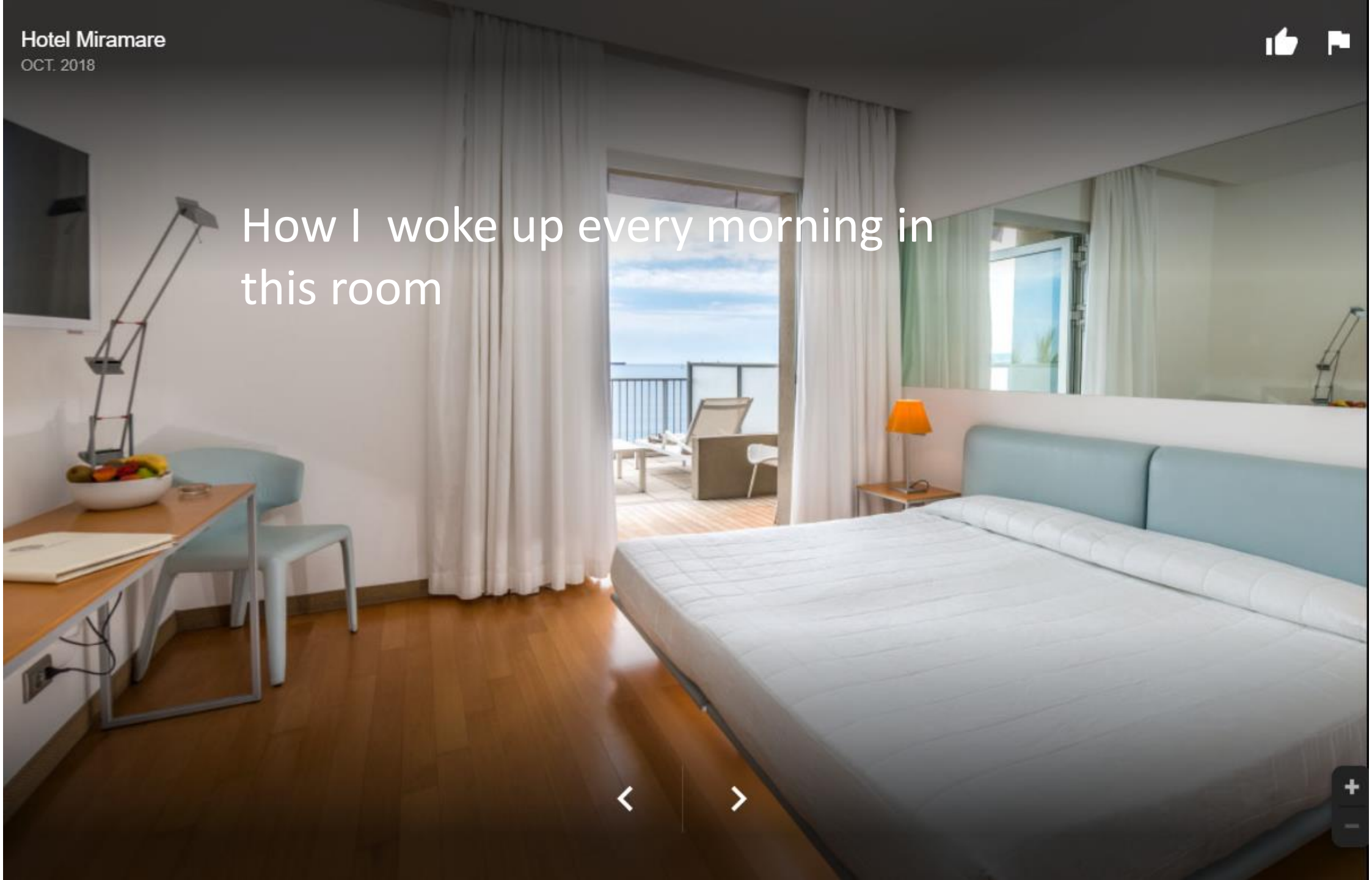
A tribute to Gianni Dal Maso

Trieste, January 31, 2020

Where I met Gianni, Sissa, 1987

Plan:

- The Mumford-Shah conjecture and our (small) contribution
- The three ways to segment an image according to Mumford-Shah: some algorithms
- A real image segmentation problem: manipulating cardboxes
- Another one : crack detection in walls
- Discussion (Bayesian formulation)


Hotel Miramare
OCT. 2018
How I woke up every morning in this room

Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems'

DAVID MUMFORD
Harvard University
AND
JAYANT SHAH
Northeastern University *(1989)*

$$E(f, \Gamma) = \mu^2 \int\int_R (f-g)^2\, dx\, dy + \int\int_{R-\Gamma} \|\nabla f\|^2\, dx\, dy + \nu|\Gamma|,$$

where $|\Gamma|$ stands for the total length of the arcs making up Γ. The smaller E is, the better (f, Γ) segments g:

(i) the first term asks that f approximates g,
(ii) the second term asks that $f-$ and hence $g-$ does not vary very much on each R_i,
(iii) the third term asks that the boundaries Γ that accomplish this be as short as possible.

Dropping any of these three items, $\inf E = 0$: without the first, take $f = 0$, $\Gamma = \varnothing$; without the second, take $f = g$, $\Gamma = \varnothing$; without the third, take Γ to be a fine grid of N horizontal and vertical lines, $R_i = N^2$ small squares, $f =$ average of g on each R_i. The presence of all three terms makes E interesting.

We do not know if the problem of minimizing E is well posed, but we conjecture this to be true. For instance we conjecture that for all continuous functions g, E has a minimum in the set of all pairs (f, Γ), f differentiable on each R_i, Γ a finite set of singular points joined by a finite set of C^1-arcs.

The second functional E_0 is simply the restriction of E to piecewise constant functions f: i.e., $f =$ constant a_i on each open set R_i. Then multiplying E by μ^{-2} we have

$$\mu^{-2}E(f, \Gamma) = \sum_i \int\int_{R_i} (g - a_i)^2 \, dx \, dy + \nu_0 |\Gamma|,$$

where $\nu_0 = \nu/\mu^2$. It is immediate that this is minimized in the variables a_i by setting

$$a_i = \text{mean}_{R_i}(g) = \int\int_{R_i} g \, dx \, dy \Big/ \text{area}(R_i),$$

so we are minimizing

$$E_0(\Gamma) = \sum_i \int\int_{R_i} (g - \text{mean}_{R_i} g)^2 \, dx \, dy + \nu_0 \cdot |\Gamma|.$$

The third functional E_∞ depends only on Γ and is given by

$$E_\infty(\Gamma) = \int_\Gamma \left[v_\infty - \left(\frac{\partial g}{\partial n} \right)^2 \right] ds,$$

where v_∞ is a constant, ds is arc length along Γ and $\partial/\partial n$ is a unit normal to Γ.

[CITATION] Boundary detection by minimizing functionals

D **Mumford**, J **Shah** - IEEE Conference on Computer Vision and ..., 1985 - San Francisco

★ 99 Cité 1050 fois Autres articles

Snakes: Active contour models | SpringerLink

de M Kass - 1988 - Cité 21510 fois - Autres articles

D. **Terzopoulos**, A. **Witkin**, and M. **Kass**, "Symmetryseeking models for 3D object reconstruction", INT. J. COMPUTER VISION, vol. 3, 1987.Google Scholar. 18.

Using these bounds, and assuming that the singular points are given by a finite number of C^2-arcs with a common endpoint, it is easy to show, by elementary comparisons of $E(f_\Gamma, \Gamma)$ with E on modified Γ's, that if $E(f_\Gamma, \Gamma)$ attains a minimum at some Γ, then the only possible singularities of Γ are:

(i) "triple points" P where three C^2-arcs meet with 120°-angles,
(ii) "crack-tips" P where a single C^2-arc ends and no other arc meets P.

Moreover, on the boundary of the domain R, another possibility is:

(iii) "boundary points" P where a single C^2-arc of Γ meets perpendicularly a smooth point of ∂R.

LEMENANT, Antoine. A selective review on Mumford–Shah minimizers. *Bollettino dell'Unione Matematica Italiana*, 2016, vol. 9, no 1, p. 69-113.

The closer result to the Mumford-Shah conjecture is probably the following one obtained by Bonnet [Bon96]. The key ingredient is a monotonicity formula for the Dirichlet integral which permitted him to classify the blow-up limits. This work is purely 2-dimensional.

Theorem 3.1. [Bon96] *Let (u, K) be a reduced minimizer. Then the Mumford-Shah conjecture is true for every isolated connected component of K. Precisely, if G is an isolated connected component of K, then it is the union of a finite set of C^1 arcs, $C^{1,1}$ away from crack-tips and that can only merge through triple junctions.*

Theorem 3.3 (Classification of connected global minimizers in $\mathbb{R}^2$). *Let (u, K) be a global minimizer in $\mathbb{R}^2$ such that K is connected. Then it belongs to the following list.*

1. *$K = \emptyset$ and u is constant.*
2. *(Line) K is a line and u is constant on each side.*
3. *(Propeller) K is the union of three half-lines meeting at their tip by angles of 120 degree.*
4. *(Cracktip) Up to translation, rotation, or additional constant, K is a half line and u is equal to the cracktip function defined in (2.6).*

[PDF] A variational method in image segmentation: existence and approximation results

G **Dal Maso**, JM Morel... - Acta Mathematica, 1992 - archive.ymsc.tsinghua.edu.cn

The main input in computer vision is the image of a scene, given by the grey level of each point of the screen. This determines a real valued measurable function g on a plane domain ff~, which, in general, is discontinuous along the lines corresponding to the edges of the ...

☆ Cité 226 fois Autres articles Les 6 versions »

Received August 7, 1989
Received in revised form June 11, 1990

[1] AMBROSIO, L., A compactness theorem for a new class of functions of bounded variation. *Boll. Un. Mat. Ital. B* (7), 3 (1989), 857–881.

[2] — Variational problems in SBV and image segmentation. *Acta Appl. Math.*, 17 (1989), 1–40.

[8] DE GIORGI, E., Free discontinuity problems in calculus of variations. *Analyse Mathématique et Applications* (*Paris*, 1988). Gauthier–Villars, Paris, 1988.

[9] DE GIORGI, E. & AMBROSIO, L., Un nuovo tipo di funzionale del calcolo delle variazioni. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.*, (1988).

[10] DE GIORGI, E., CARRIERO, M. & LEACI, A., Existence theorem for a minimum problem with free discontinuity set. *Arch. Rational Mech. Anal.*, 108 (1989), 195–218.

LEMENANT, Antoine. A selective review on Mumford–Shah minimizers. *Bollettino dell'Unione Matematica Italiana*, 2016, vol. 9, no 1, p. 69-113.

3.2.2 K has the uniform concentration property.

The uniform concentration property is a very important case of lower semisontinuity introduced by Dal Maso, Morel and Solimini [DMMS92] which allows to take limits of minimizers and prove that the limit is still a minimizer. Indeed, it is very classical that $K \mapsto \mathcal{H}^{N-1}(K)$ may not be lower semicoutinuous with respect to Hausdorff convergence in general. A very famous special case in dimension 2 is the so called Golab Theorem which says that it holds true along a sequence of compact connected sets. The uniform concentration property provides a sort of generalization which works in any dimension, and says that $K \mapsto \mathcal{H}^{N-1}(K)$ is lower semicontinuous for a uniform sequence of uniformly concentrated sets, which actually holds for a sequence of Mumford-Shah minimizers.

Definition 0.9. Let B be a Borel subset of $\bar{\Omega}$. We say that B satisfies the *concentration property* in Ω if for every $\varepsilon>0$ there exists $\alpha=\alpha(\varepsilon)>0$ such that, if $D_R=D(x_0,R)$ is any disc contained in Ω with $x_0\in B$ and $0<R<1$, then there exists a disc $D=D(x,r)$ contained in D_R such that

$$\operatorname{diam}(D)\geqslant \alpha \operatorname{diam}(D_R),$$

$$\mathcal{H}^1(D\cap B)\geqslant (1-\varepsilon)\operatorname{diam}(D).$$

Roughly speaking, this property says that any disc centered on B contains a subdisc, with comparable diameter, where B is concentrated.

To obtain our inequality (0.8) we use the following lower semicontinuity result.

LEMMA 0.10 (Lower Semicontinuity Lemma). *Let (K_k) be a sequence of closed subsets of Ω which converges in the Hausdorff metric to a closed subset K of $\bar{\Omega}$. Assume that the sets K_k satisfy the concentration property in Ω (Definition* 0.4*) uniformly with respect to k (i.e. with $\alpha(\varepsilon)$ independent of k). Then*

$$\mathcal{H}^1(K\cap\Omega)\leqslant \liminf_{k\to\infty} \mathcal{H}^1(K_k\cap\Omega). \tag{0.11}$$

LEMENANT, Antoine. A selective review on Mumford–Shah minimizers. *Bollettino dell'Unione Matematica Italiana*, 2016, vol. 9, no 1, p. 69-113.

THEOREM 0.5 (Convergence Theorem). *For every* $k \in \mathbf{N}$ *let* $(u_k; \gamma_k^1, \ldots, \gamma_k^k)$ *be a minimum point for* (0.3). *Assume that the sets*

$$K_k = \bigcup_{i=1}^{k} \gamma_k^i([0,1])$$

have no isolated points. Then there exists a subsequence of (u_k, K_k) *which converges to a minimum point* (u, K) *of* (0.1) *in the following sense:*

(a) $K_k \to K$ *in the Hausdorff metric,*

(b) $u_k \to u$ *strongly in* $L^2(\Omega)$,

(c) $J_k(u_k; \gamma_k^1, \ldots, \gamma_k^k) \to J(u, K)$.

“Short list” of authors who dedicated time to the Mumford-Shah conjecture:

Giovanni Alberti, Guy Bouchitté, Gianni Dal Maso, Luigi Ambrosio, Nicola Fusco, John E. Hutchinson, Diego Pallara, Vincenzo Maria Tortorelli, Giovanni Bellettini, Alessandra Coscia, Blaise Bourdin, Antonin Chambolle, Jean-François Babadjian, Antoine Lemenant, Alexis Bonnet, Guy David, Dorin Bucur, Stephan Luckhaus, Gilles A. Francfort, Jean-Jacques Marigo, Ennio De Giorgi, Michele Carriero, Antonio Leaci, Françoise Dibos, Georges Koepfler, Jean-Christophe Léger, Camillo De Lellis, Matteo Focardi, Berardo Ruffini, Rodica Toader, Sergio Solimini, Guido De Philippis, Alessio Figalli, Thierry De Pauw, Didier Smets, Stephen Semmes, Nicola Fusco Matteo Focardi, Berardo Ruffini, Benoît Merlet, Maria Giovanna Mora, Massimiliano Morini, Séverine Rigot, Anthony Siaudeau.

What was done to tackle (define, solve) the “real” image segmentation problem?

The three ways to segment an image according to Mumford-Shah: some algorithms

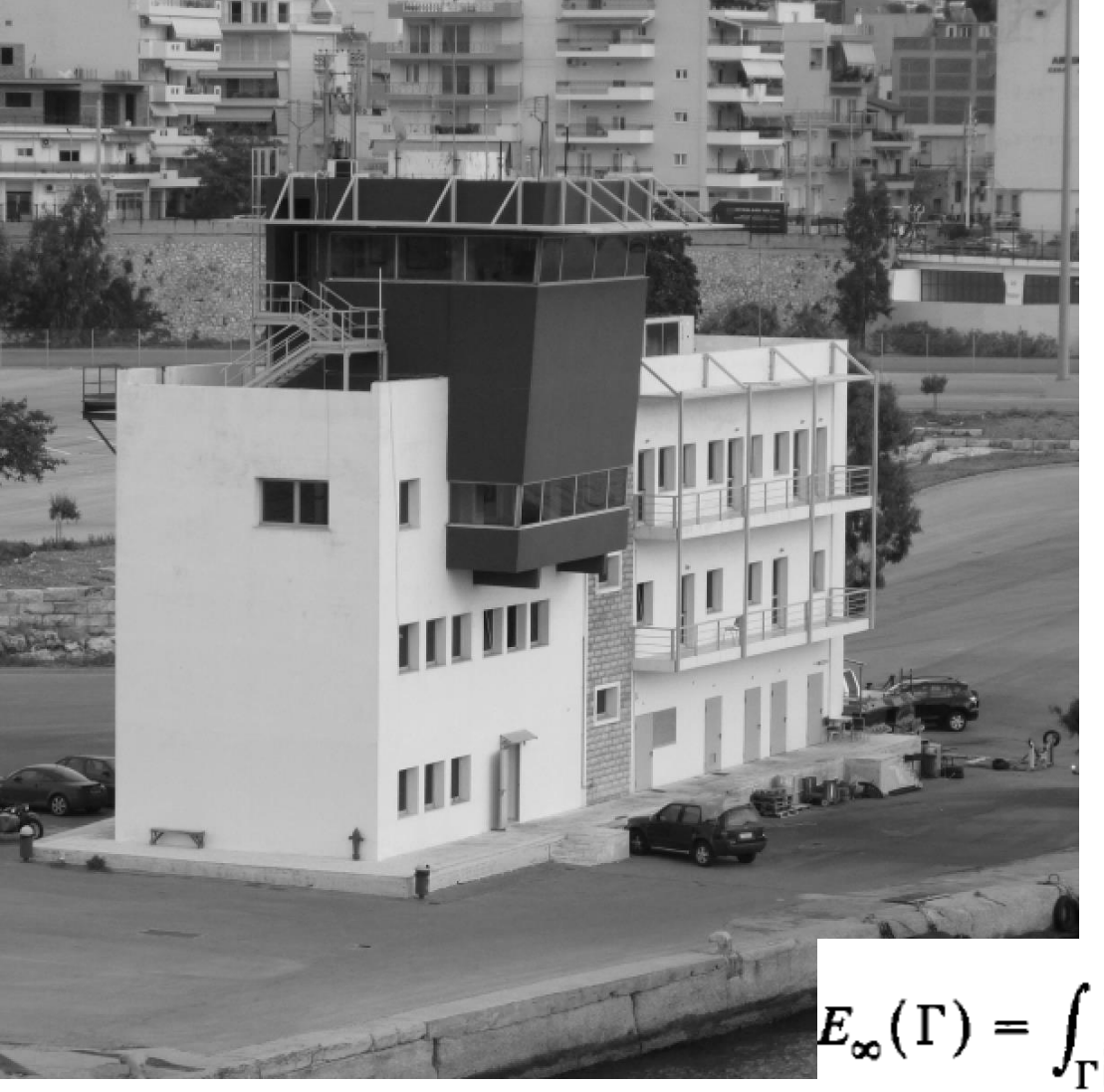
The functional works pretty well in practice. A numerical method can be obtained using the phase-field approximation of Ambrosio-Tortorelli [AT92, Bou99, BC94] or even by a direct finite elements method [CDM99, BC00].

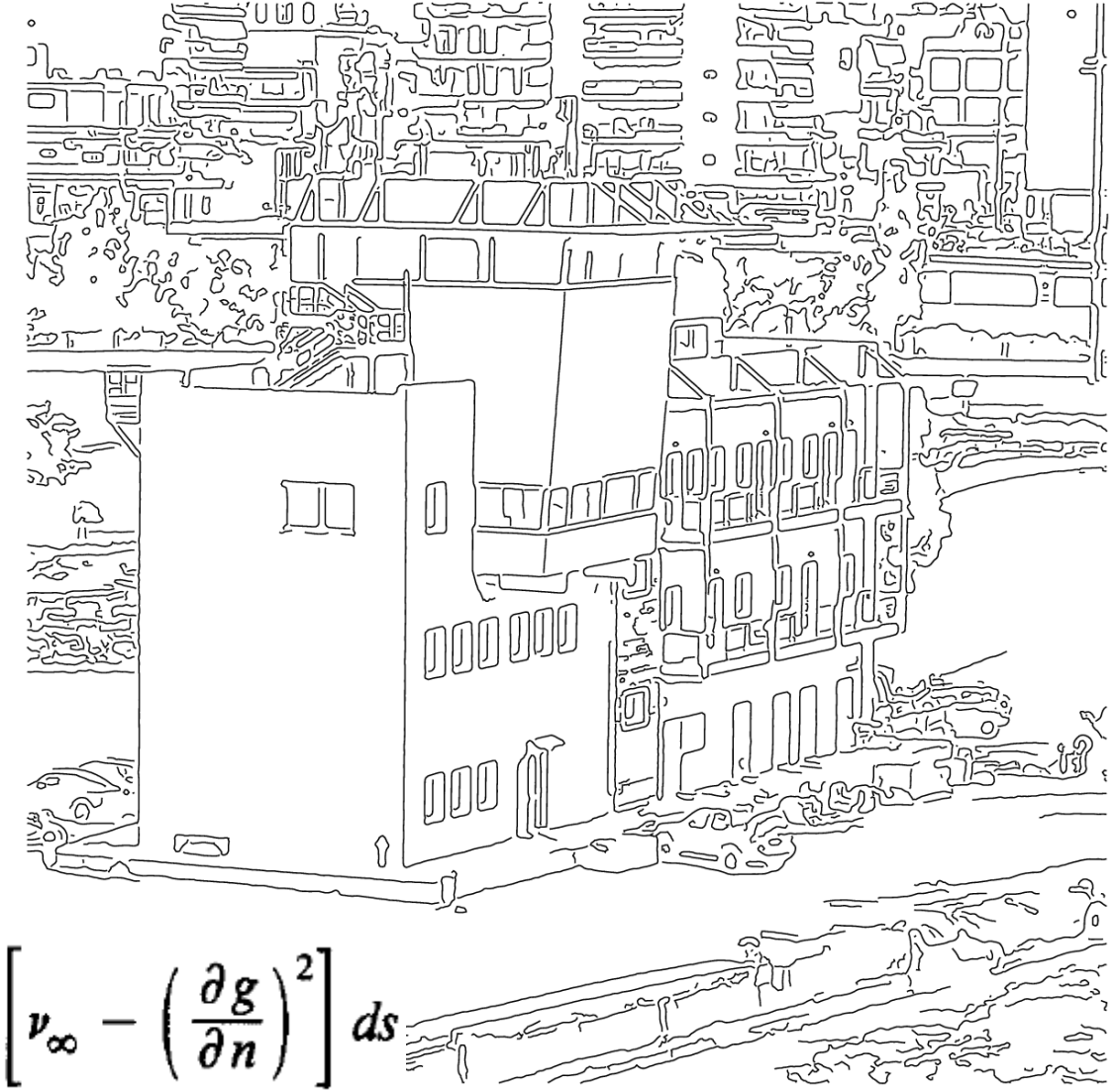
Segmentation of a Cagou by use of the Mumford-Shah functional.

Antoine Lemenant. A selective review on Mumford-Shah minimizers. Bollettino dell'Unione Matematica Italiana, Springer Verlag, 2016, 9, pp.69 - 113.

$$E_{\infty}(\Gamma) = \int_{\Gamma}\left[\nu_{\infty} - \left(\frac{\partial g}{\partial n}\right)^2\right] ds$$

VON GIOI, Rafael Grompone et RANDALL, Gregory. A Sub-Pixel Edge Detector: an Implementation of the Canny/Devernay Algorithm. *IPOL Journal*, 2017, vol. 7, p. 347-372.





$$E_\infty(\Gamma) = \int_\Gamma \left[\nu_\infty - \left(\frac{\partial g}{\partial n} \right)^2 \right] ds$$

VON GIOI, Rafael Grompone et RANDALL, Gregory. A Sub-Pixel Edge Detector: an Implementation of the Canny/Devernay Algorithm. *IPOL Journal*, 2017, vol. 7, p. 347-372.

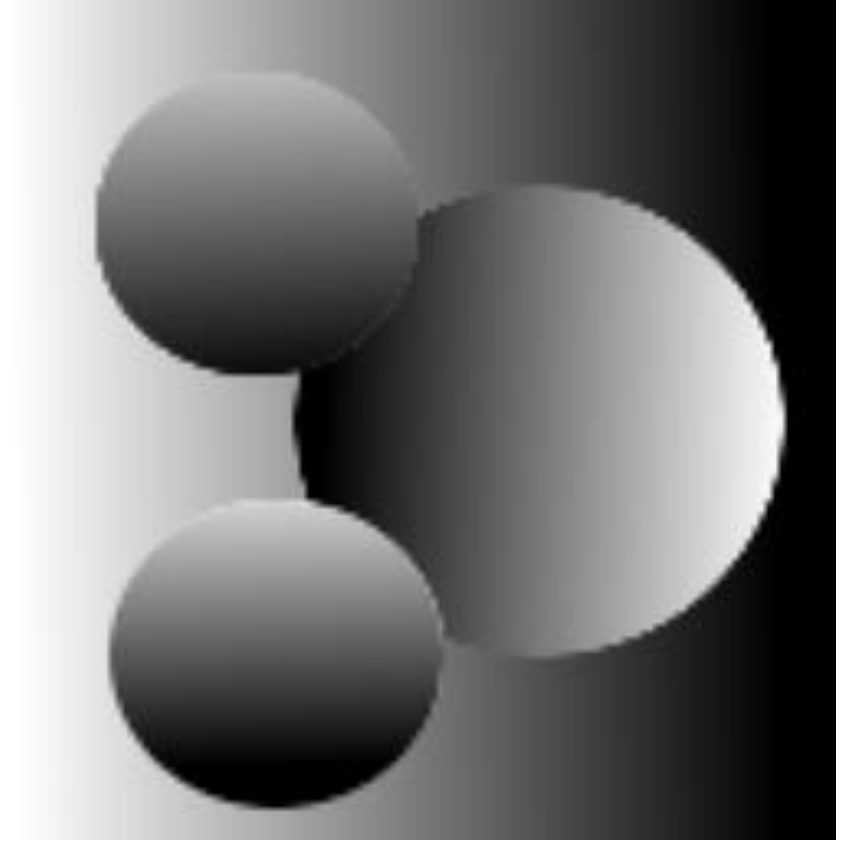
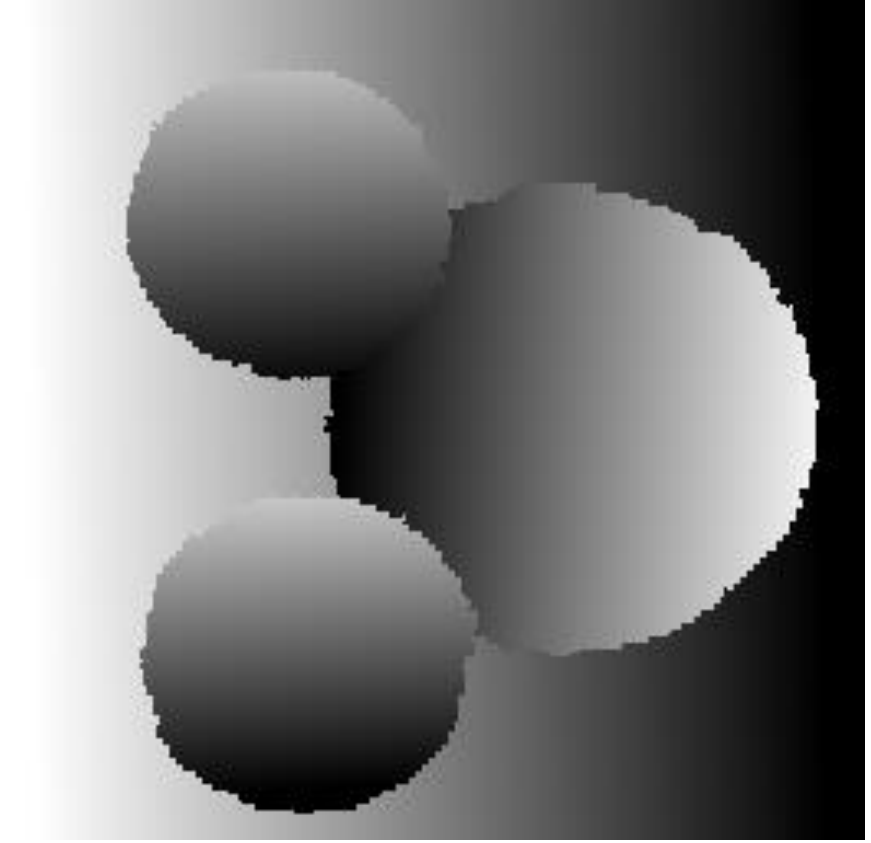
Von Gioi, R. G., Jakubowicz, J., M., J. M., & Randall, G. (2012):
LSD: a line segment detector. *Image Processing On Line*, *2*, 35-55.

Von Gioi, R. G, Jakubowicz, J., M., J. M., & Randall, G. (2012). LSD: a line segment detector. *Image Processing On Line*, *2*, 35-55.

$$E_n(\{R_i\}) = \sum_i \left(\iint (g - \text{mean}_{R_i} g)^2 \, dx \, dy + \tfrac{1}{2} \nu_0 |\partial R_i| \right).$$

Piecewise Affine Image Segmentation Based on Mumford-Shah Functional (4 regions) G. Facciolo

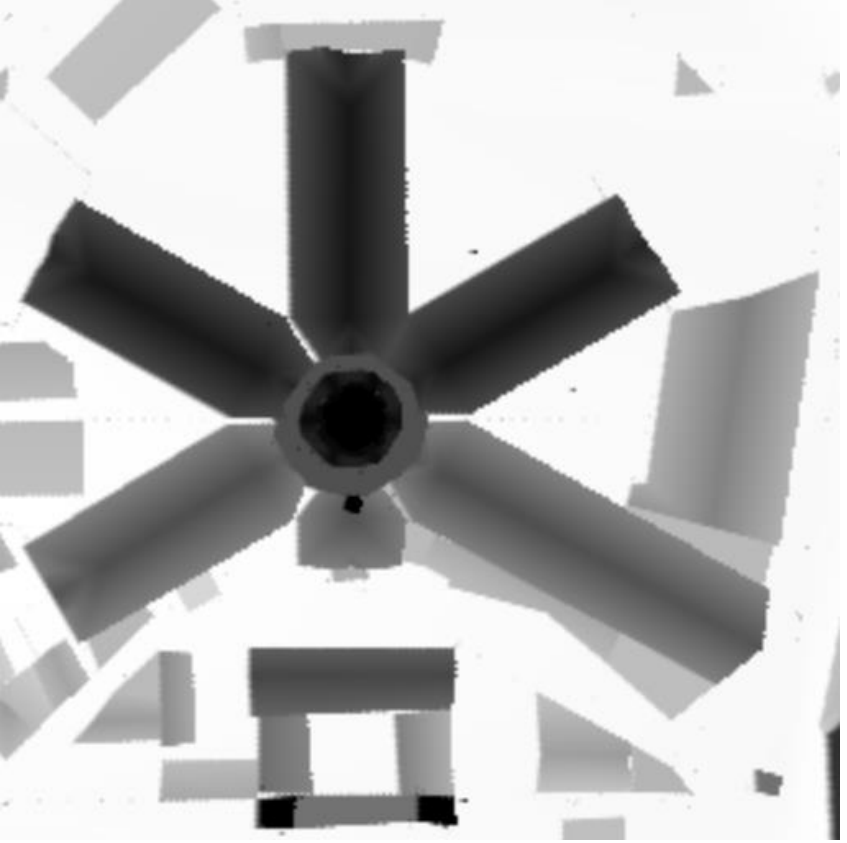
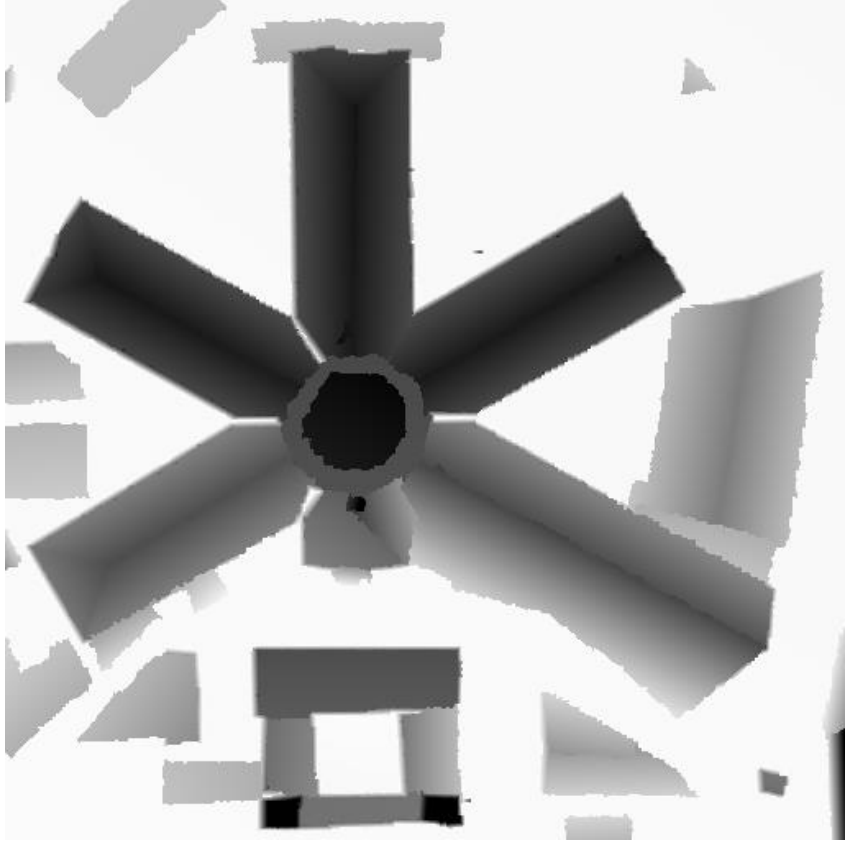
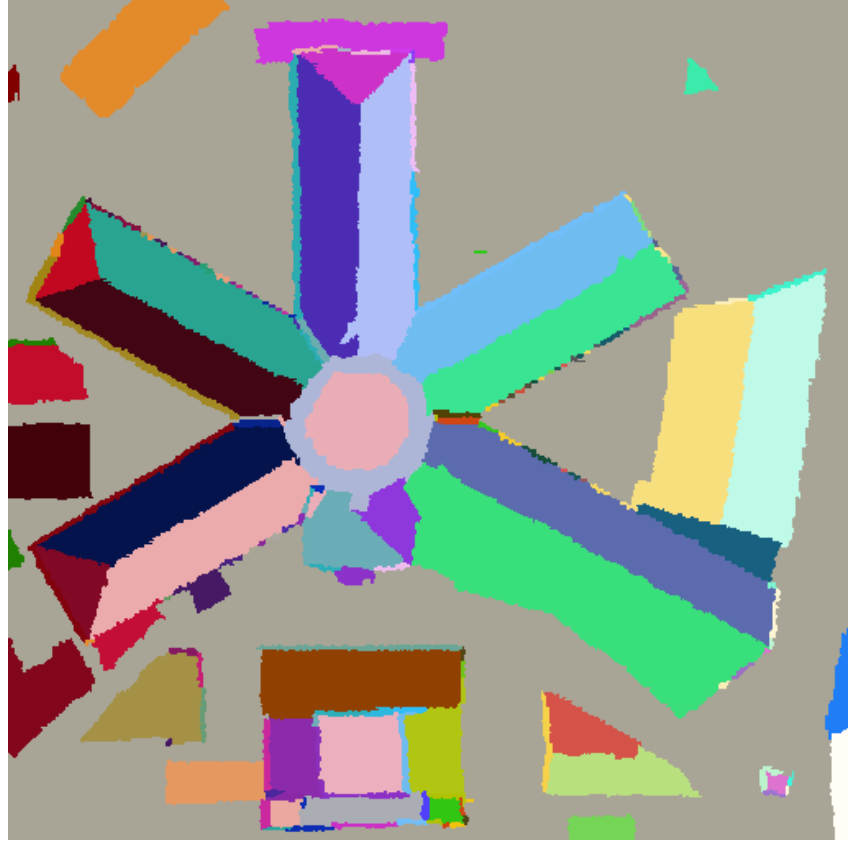
Algorithm: Koepfler, G., Lopez, C., & Morel, J. M. (1994). A multiscale algorithm for image segmentation by variational method. *SIAM journal on numerical analysis*, *31*(1), 282-299.


$$E_n(\{R_i\}) = \sum_i \left(\iint (g - \text{mean}_{R_i} g)^2 \, dx \, dy + \tfrac{1}{2} \nu_0 |\partial R_i| \right).$$

Piecewise Affine Image Segmentation Based on Mumford-Shah Functional (4 regions)
G. Facciolo

Algorithm: Koepfler, G., Lopez, C., & Morel, J. M. (1994). A multiscale algorithm for image segmentation by variational method. *SIAM journal on numerical analysis*, *31*(1), 282-299.

$$E_n(\{R_i\}) = \sum_i \left(\iint (g - \text{mean}_{R_i} g)^2 \, dx \, dy + \tfrac{1}{2} v_0 |\partial R_i| \right).$$

Piecewise Affine Image Segmentation Based on Mumford-Shah Functional (4 regions)
G. Facciolo

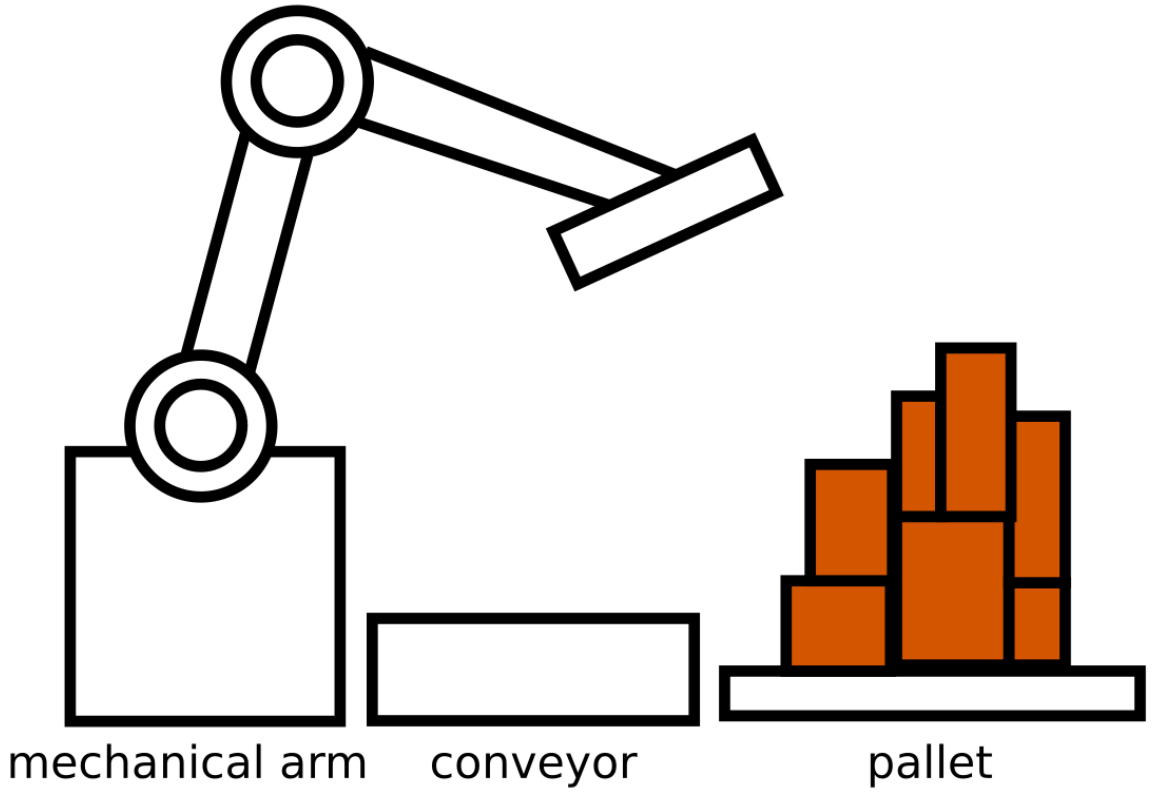
Algorithm: Koepfler, G., Lopez, C., & Morel, J. M. (1994). A multiscale algorithm for image segmentation by variational method. *SIAM journal on numerical analysis*, *31*(1), 282-299.

A jump in time: A real segmentation problem in 2020: Cardbox detection

(joint work with Sébastien Drouyer)

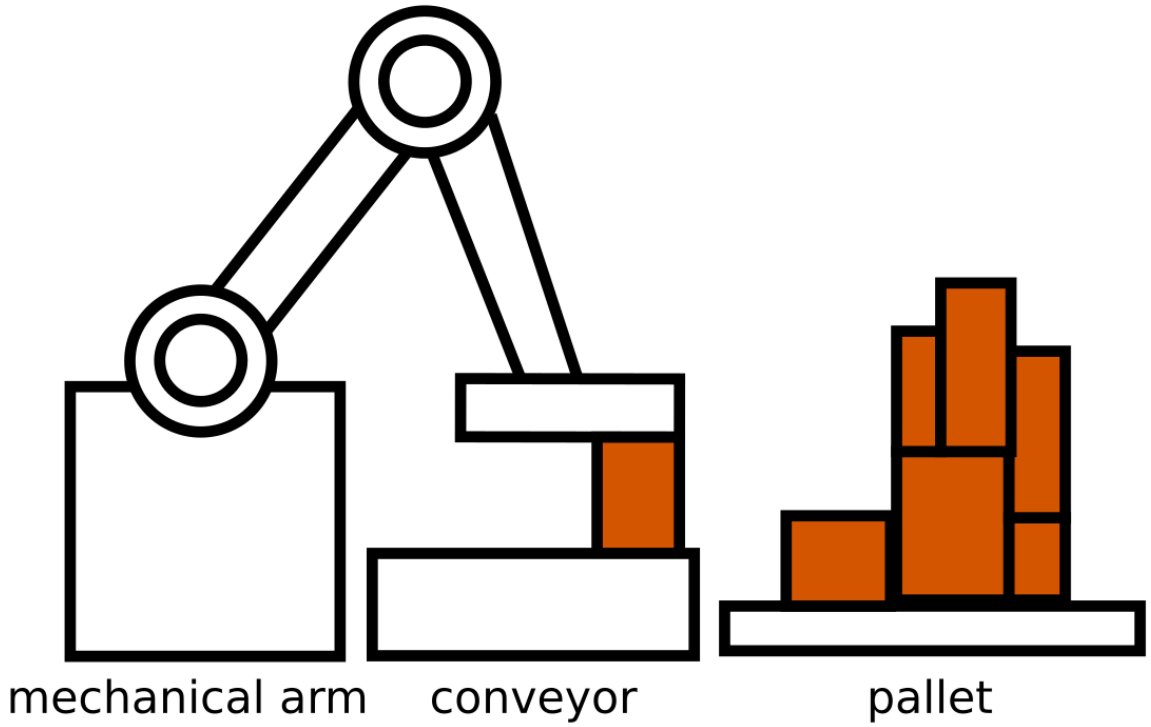
Cardbox detection

Due to labor shortage, warehouses in Japan need to automate their



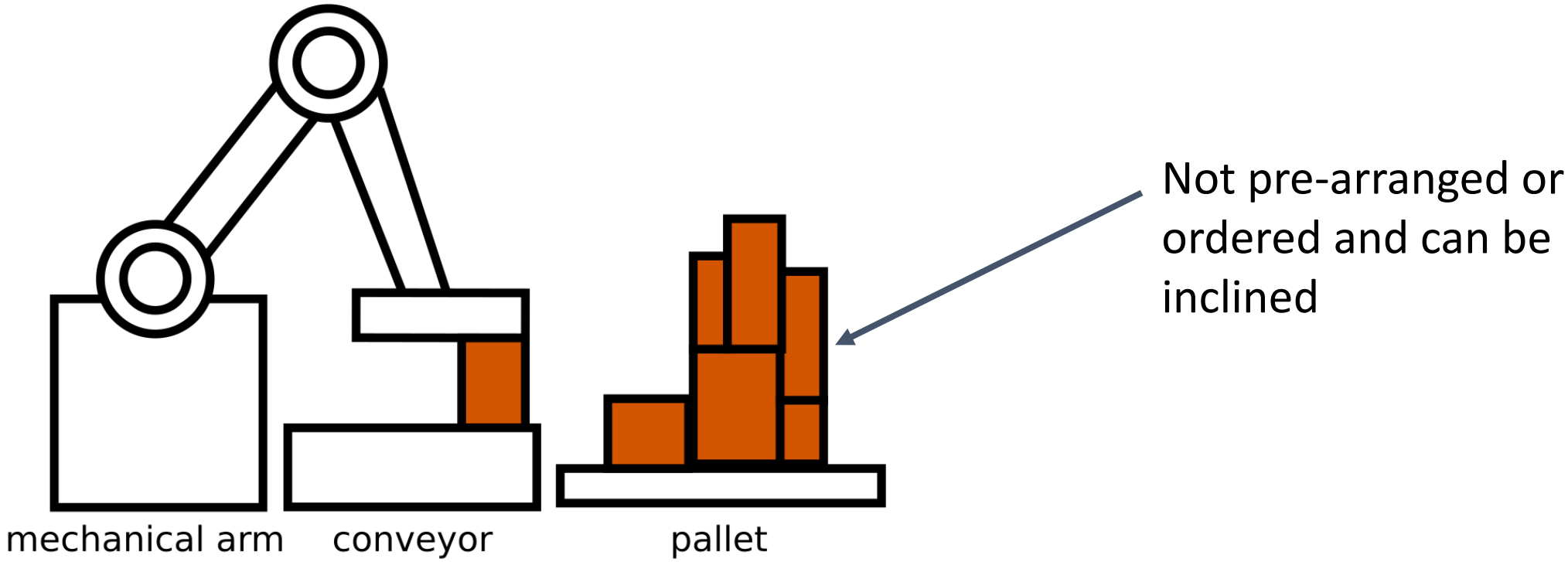
Cardbox detection

Due to labor shortage, warehouses in Japan need to automate their



Cardbox detection

Due to labor shortage, warehouses in Japan need to automate their



Cardbox detection

Input 3D + 2D (so far)

Cardbox detection

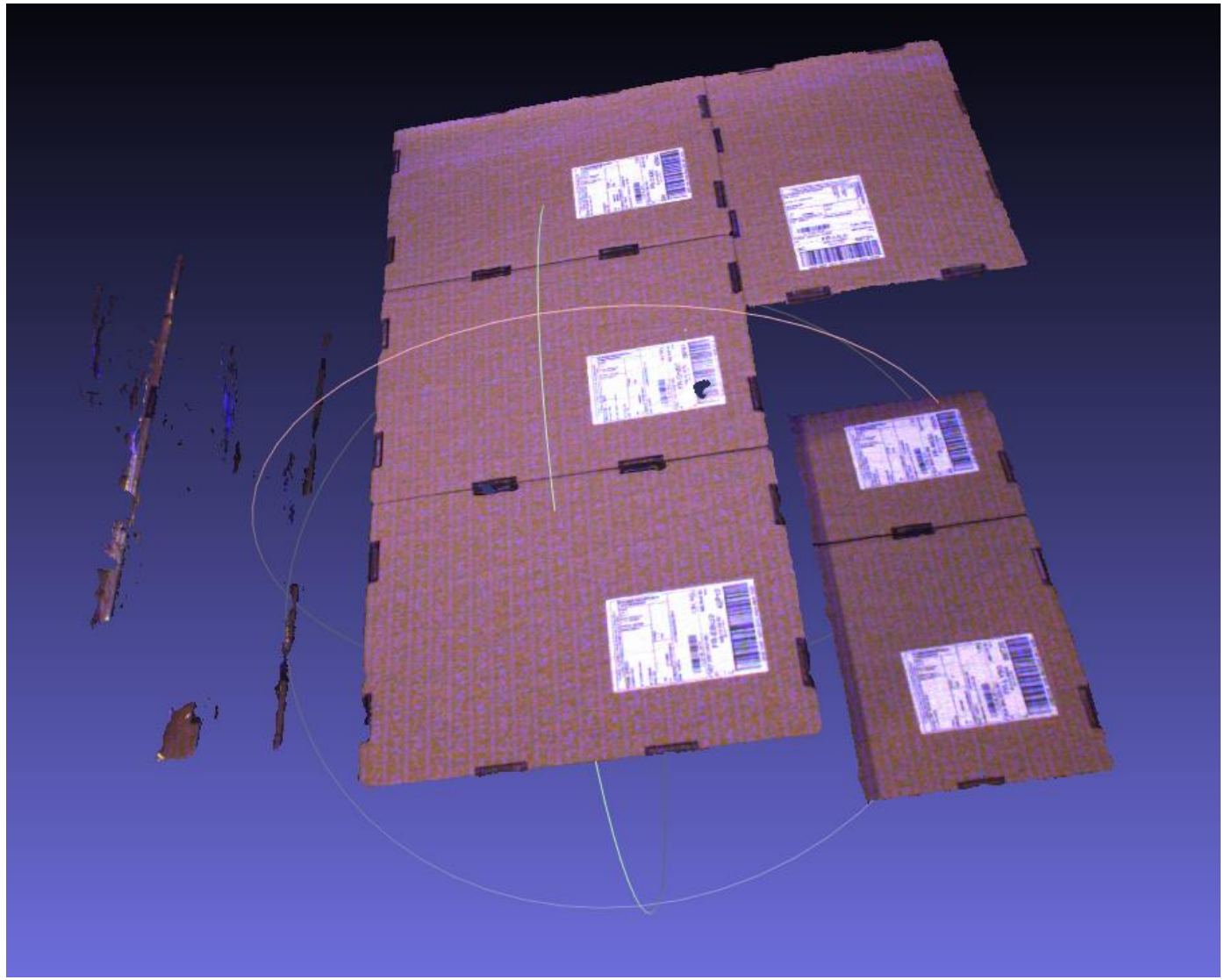
Input 3D + 2D (so far)

Cardbox detection

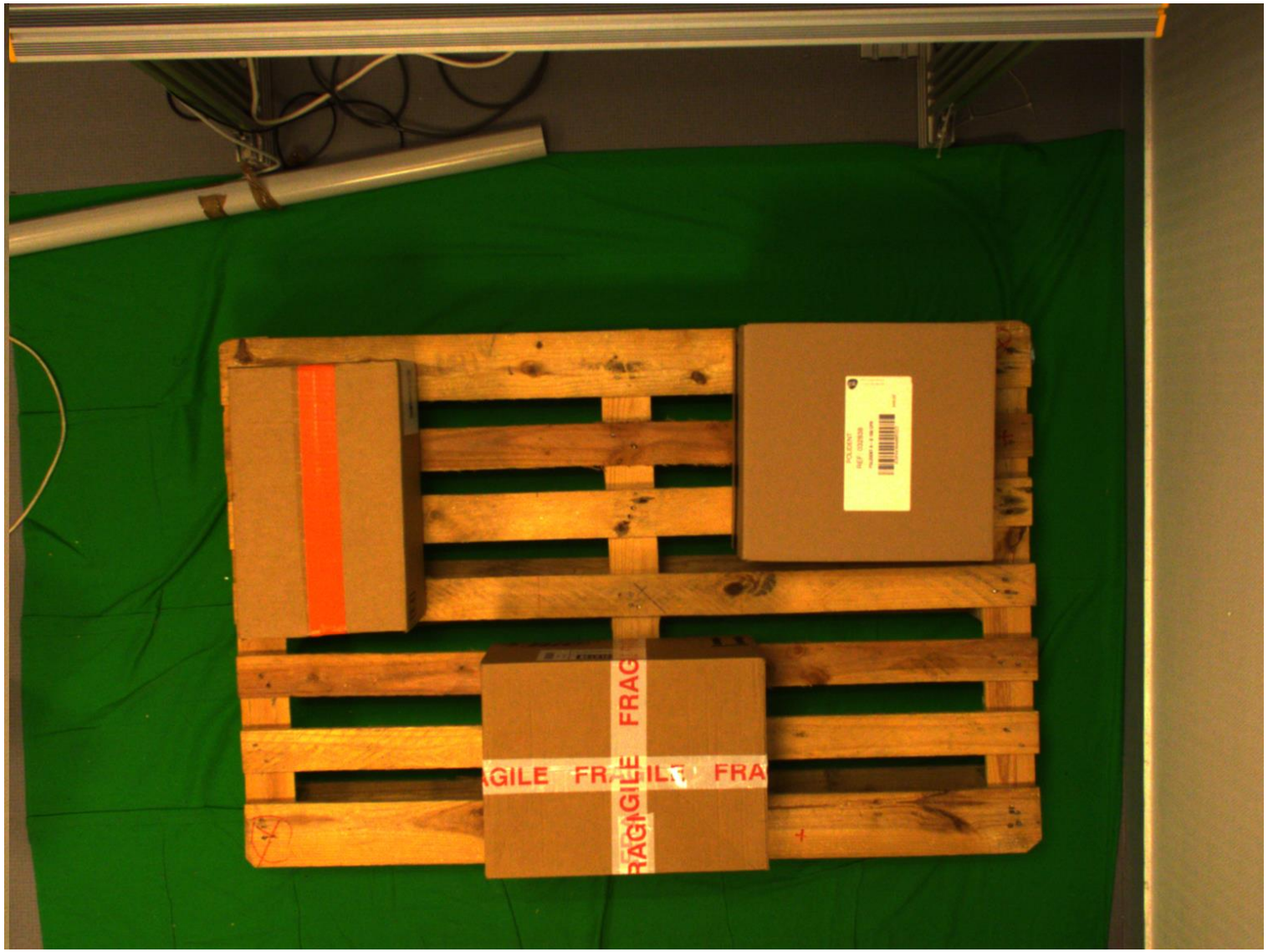
Input 3D + 2D (so far)

Cardbox detection

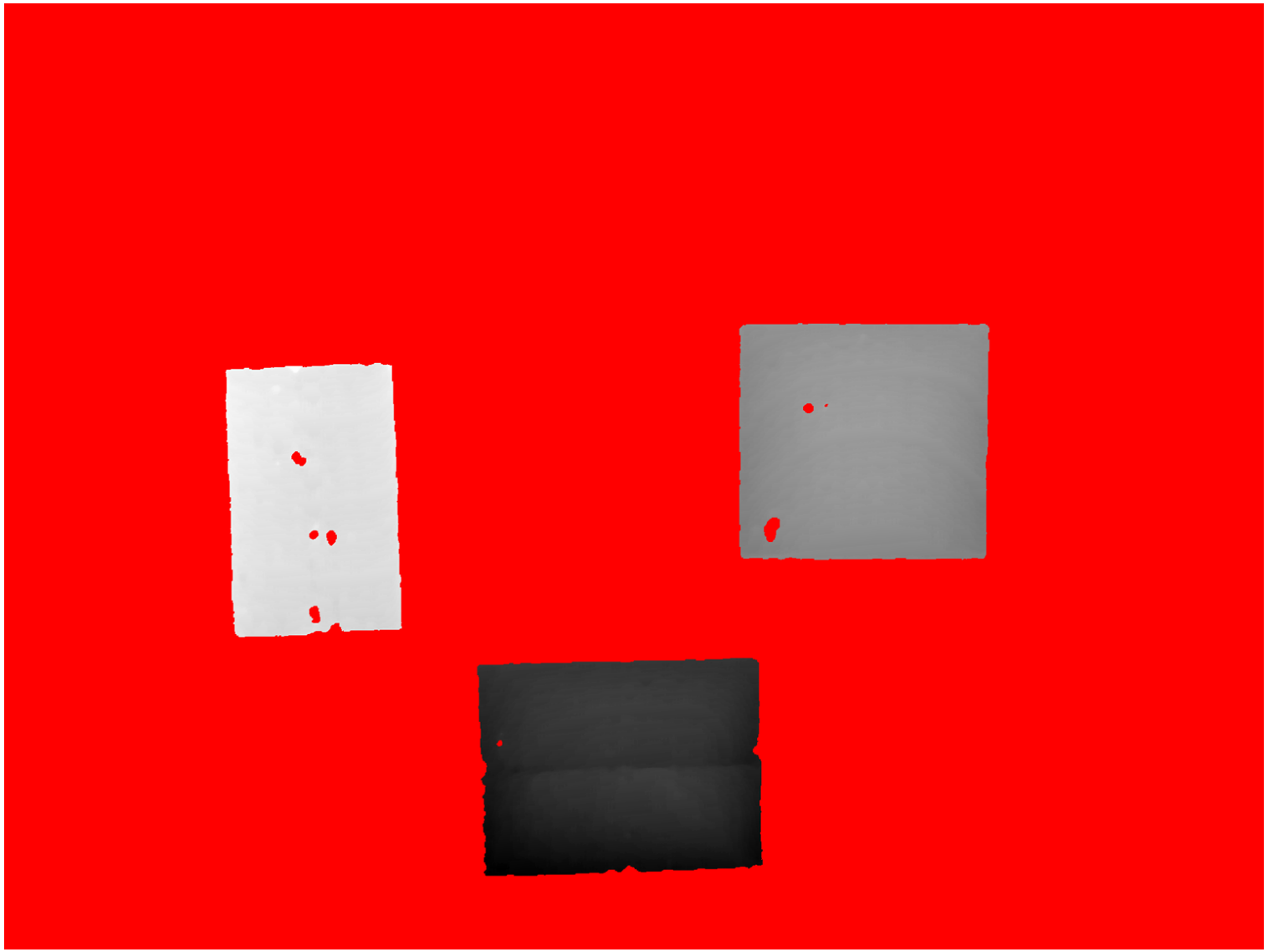
Input 3D + 2D (so far)



Cardbox detection

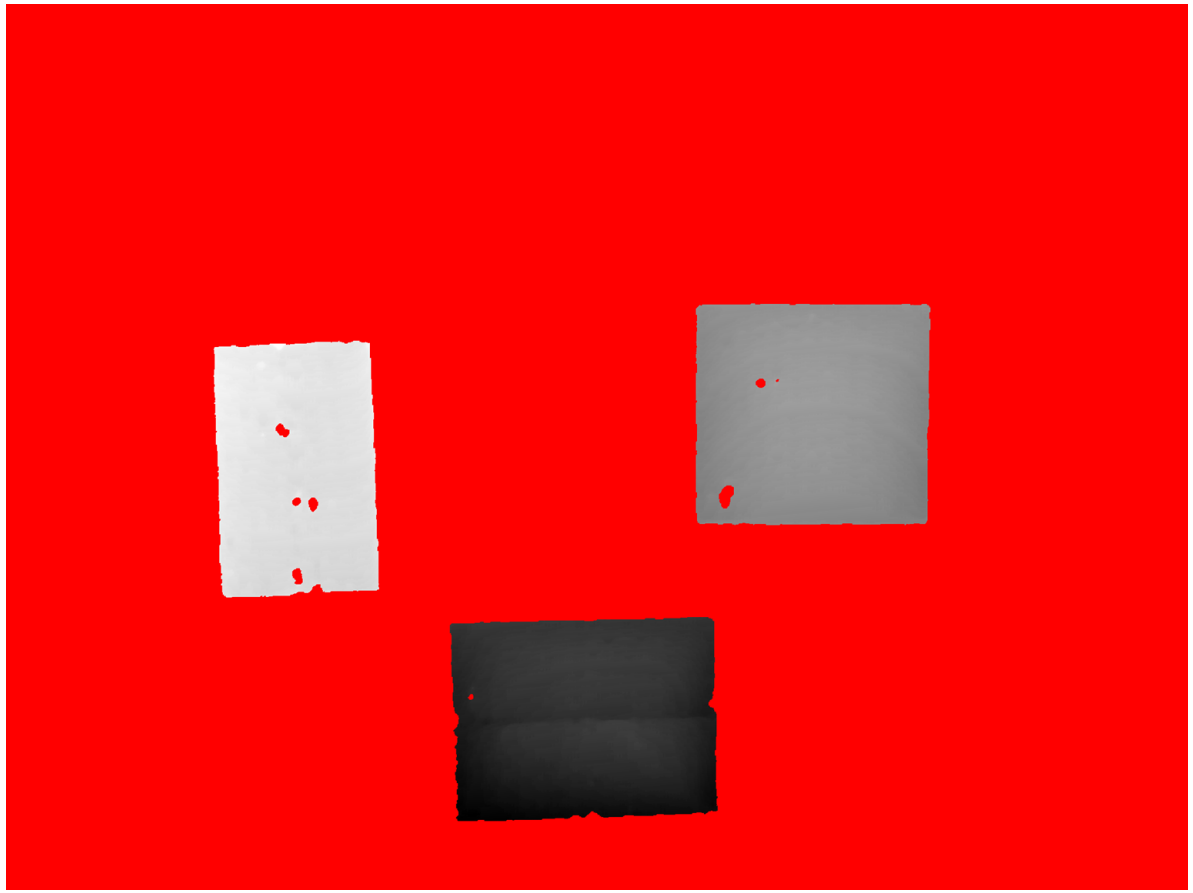


Cardbox detection



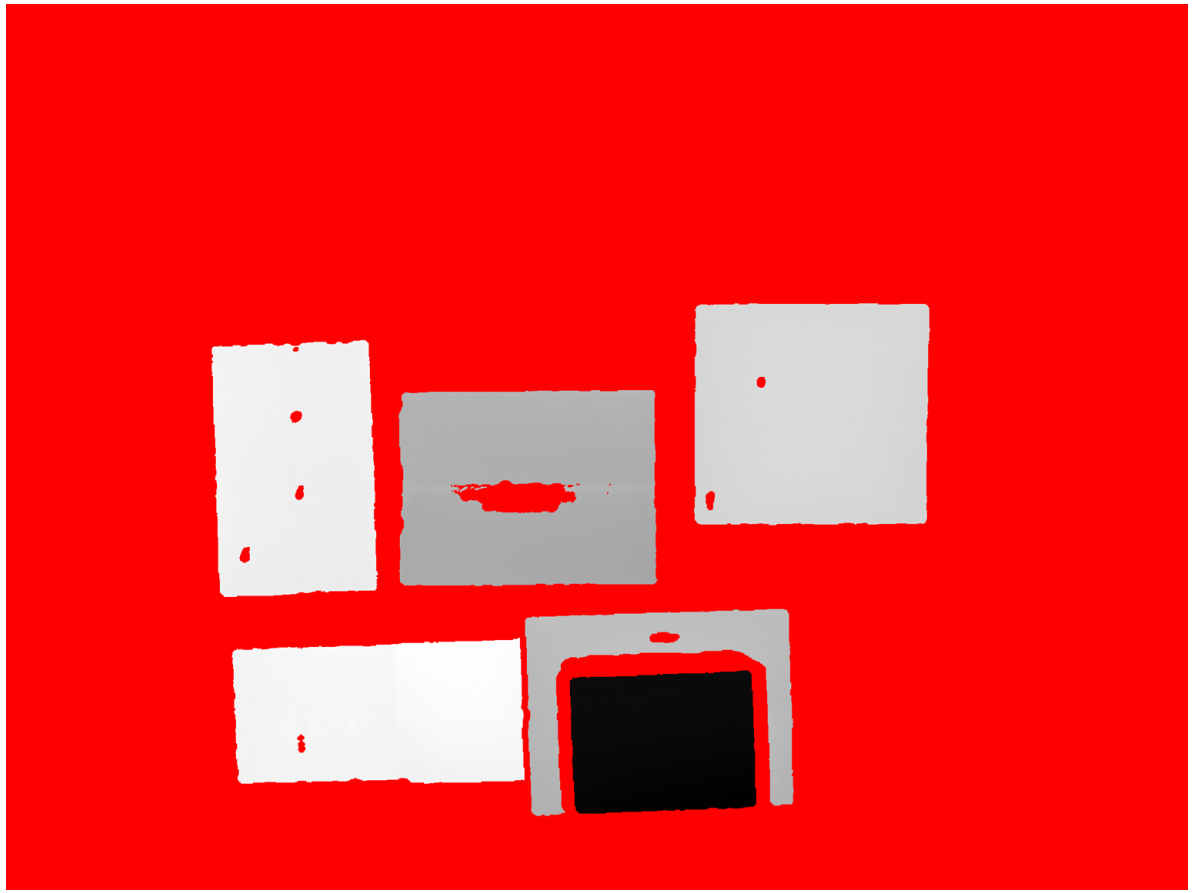
Cardbox detection

Most times packages can easily be separated using the 3D map.



Cardbox detection

But not always.

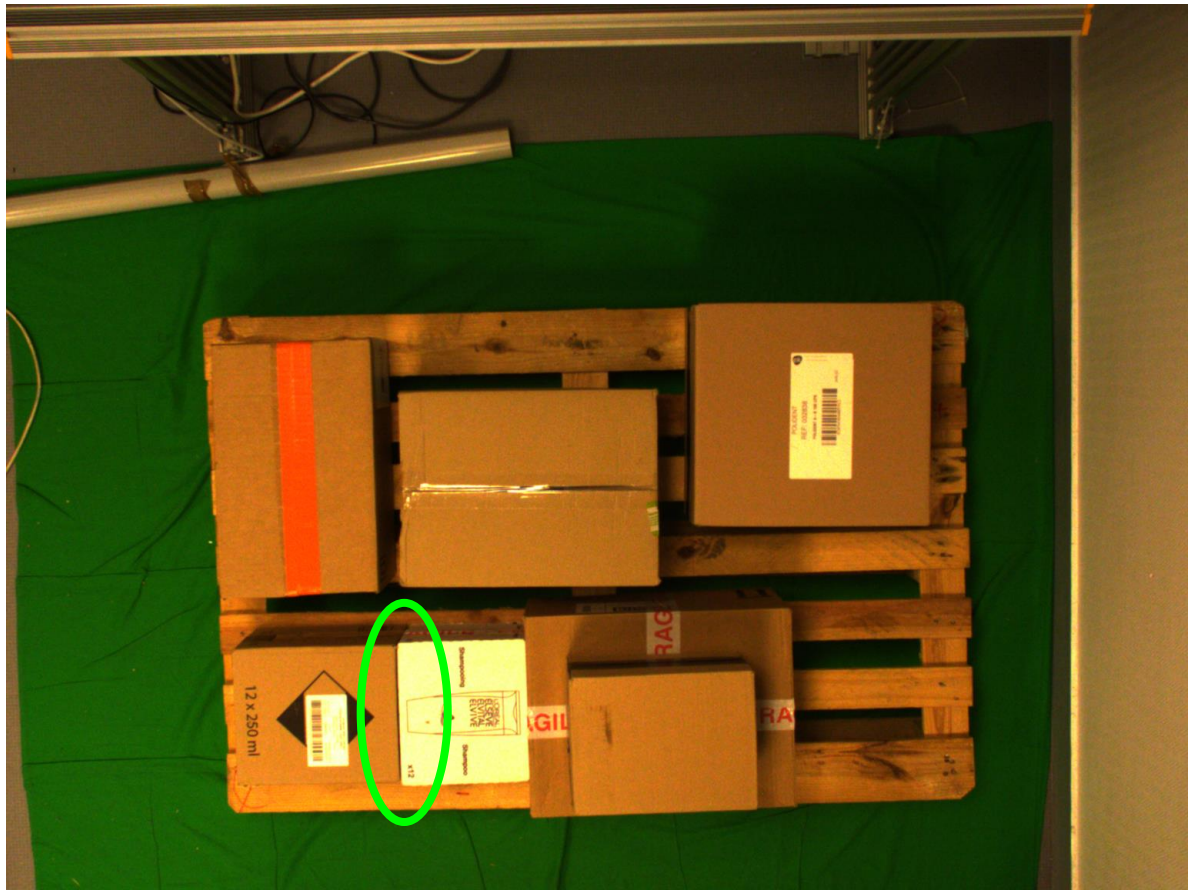


Cardbox detection

But not always.

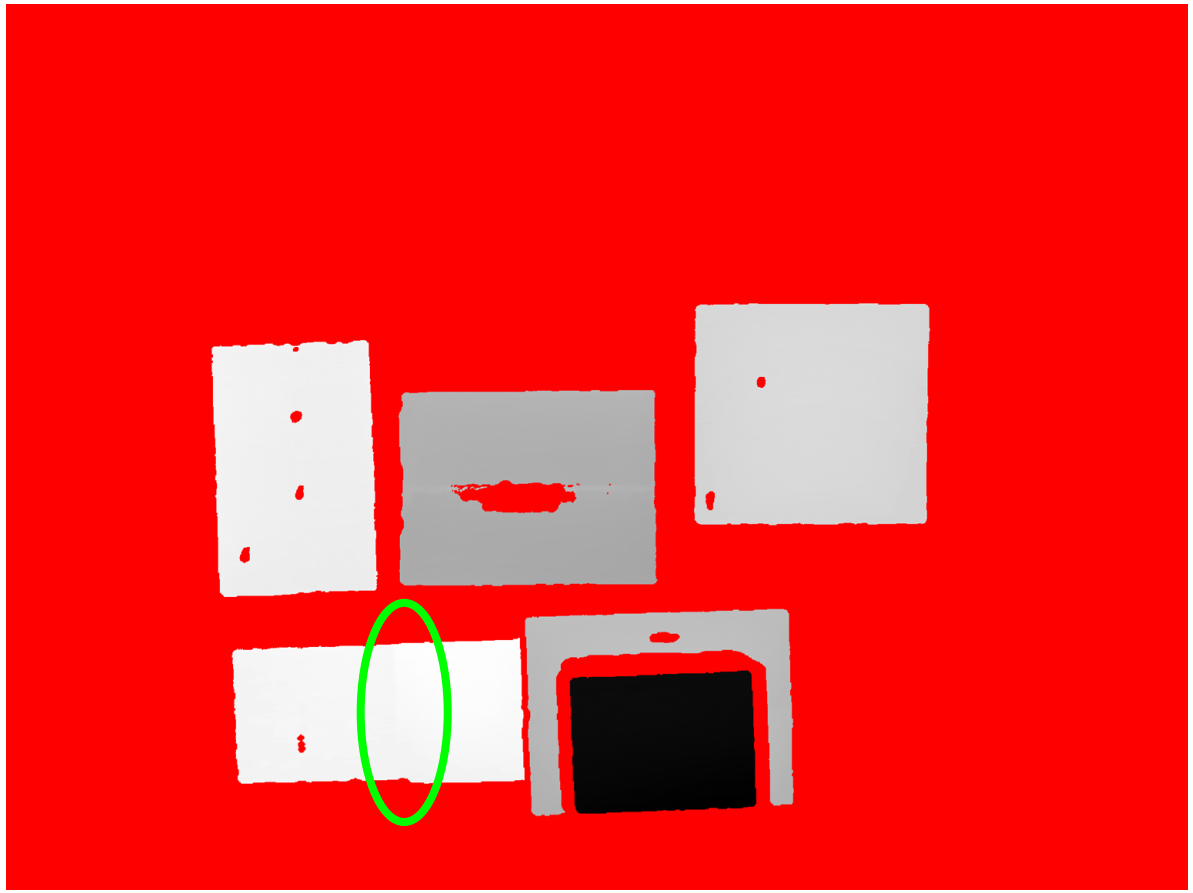
Cardbox detection

But not always.



Cardbox detection

But not always.



Cardbox detection

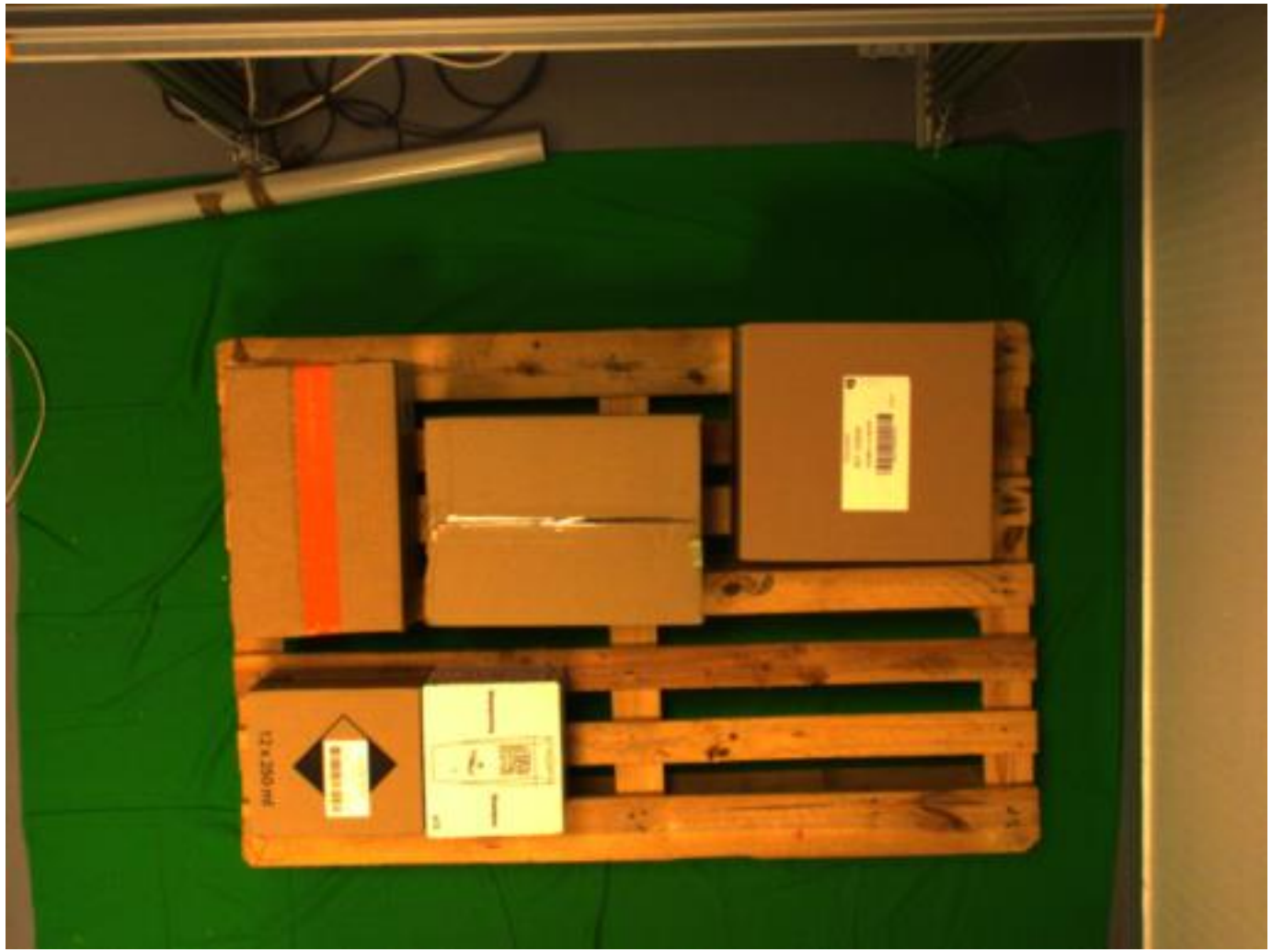
→ Deep learning

Main idea:

- Use a first segmentation obtained by 3D
- Refine this segmentation using semantic segmentation algorithm on 2D

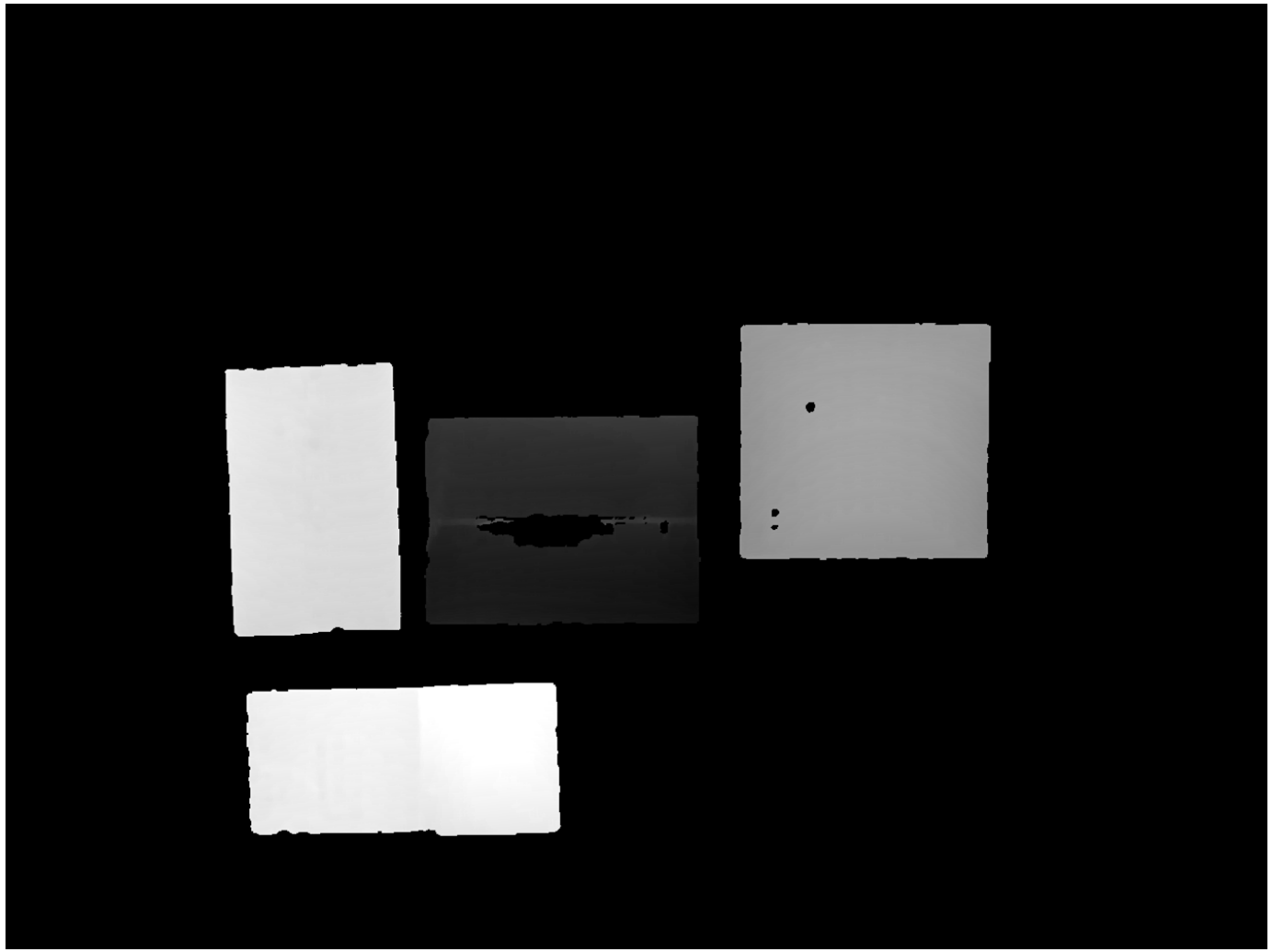
Cardbox detection

Image



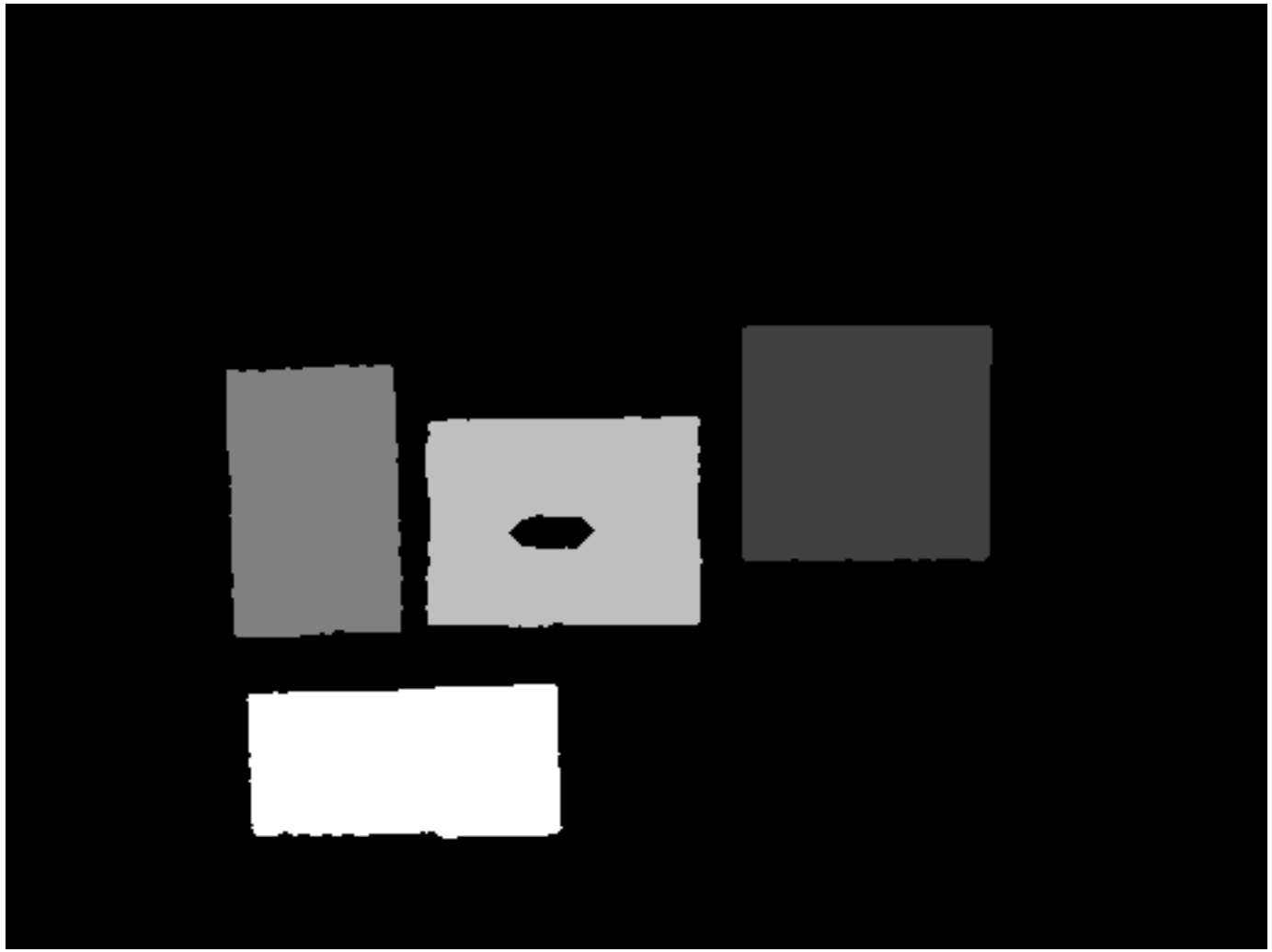
Cardbox detection

3D



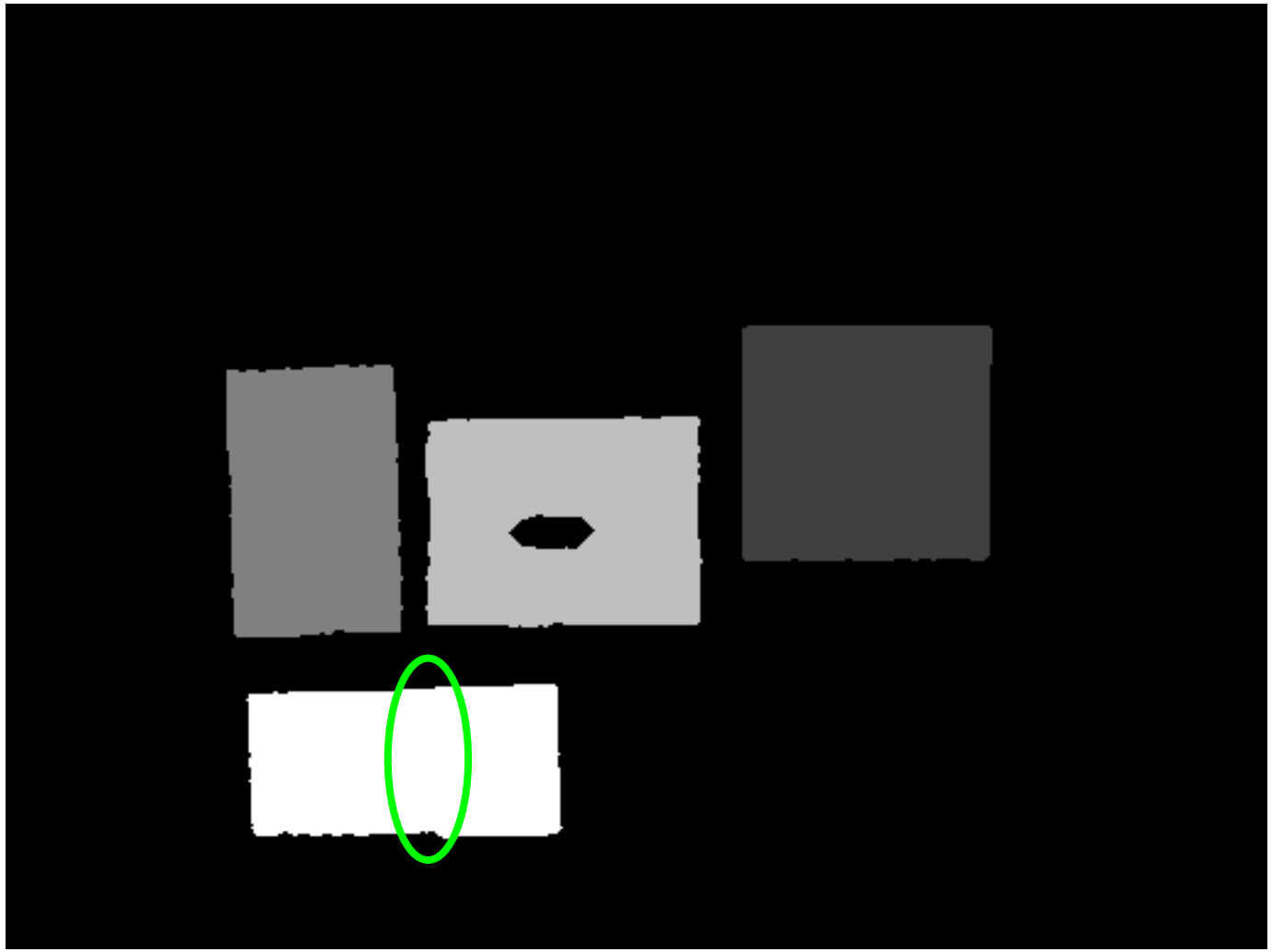
Cardbox detection

Label of 3D



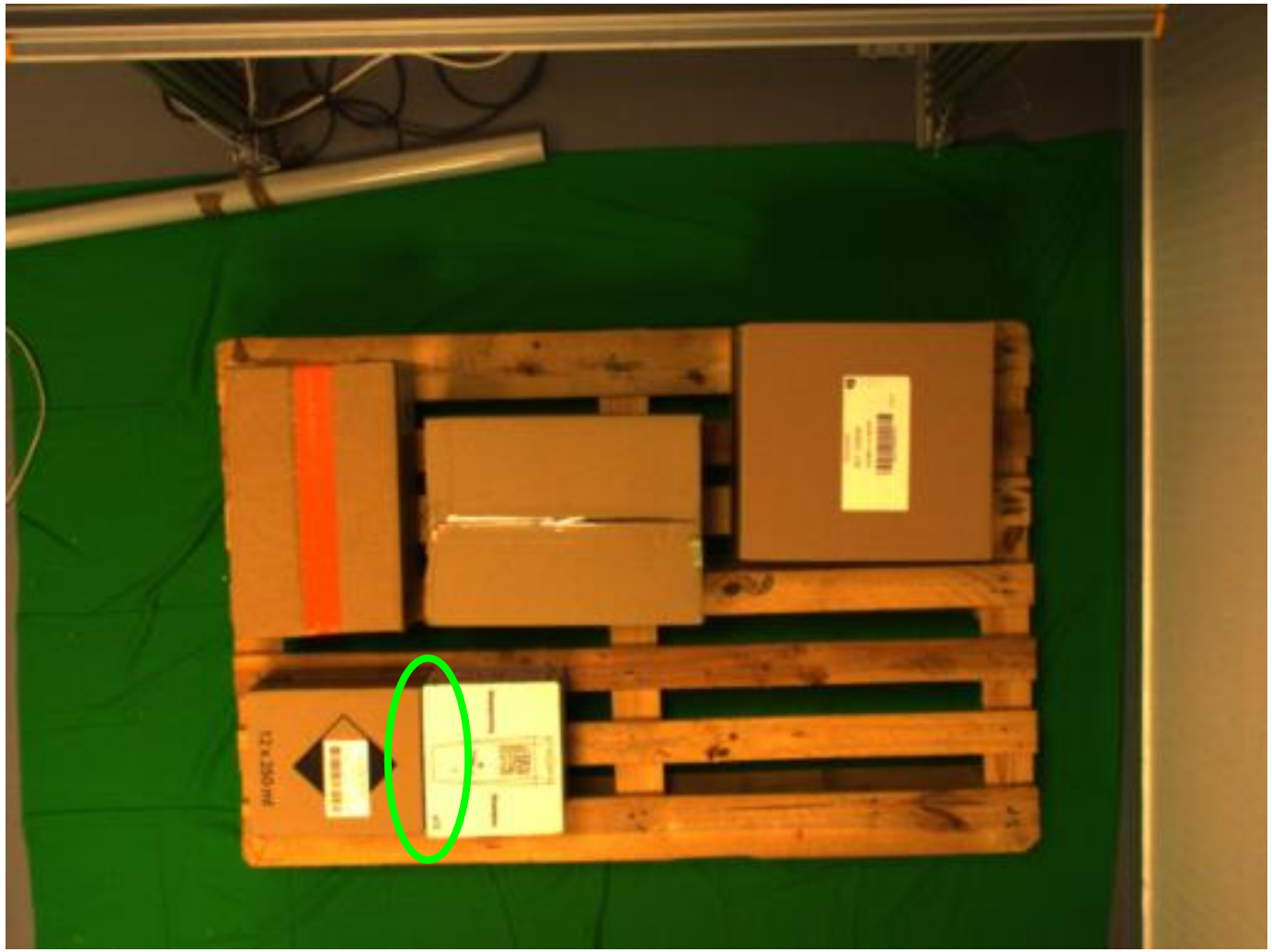
Cardbox detection

Label of 3D



Cardbox detection

One « semantic boundary » is missing

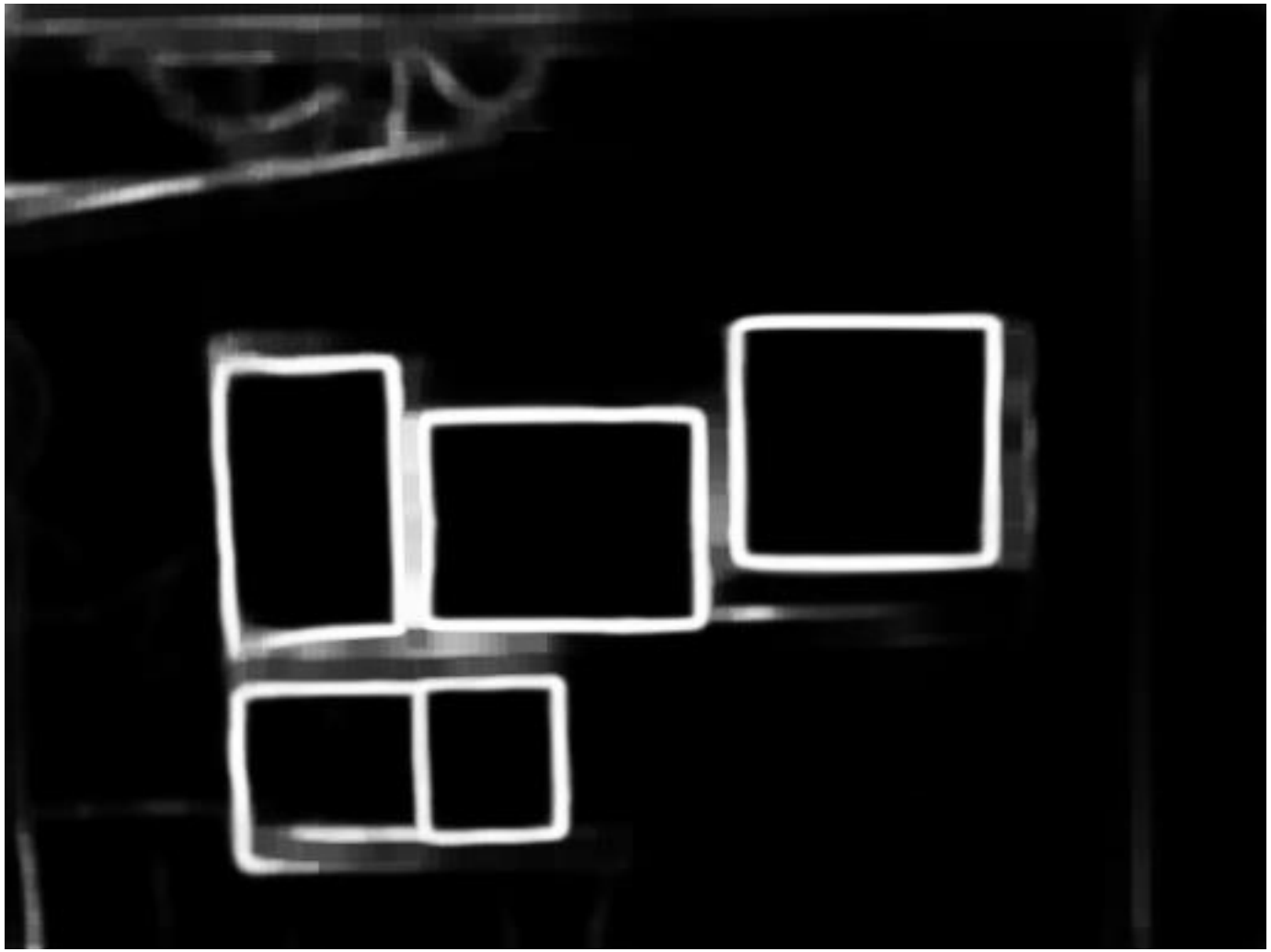


Cardbox detection

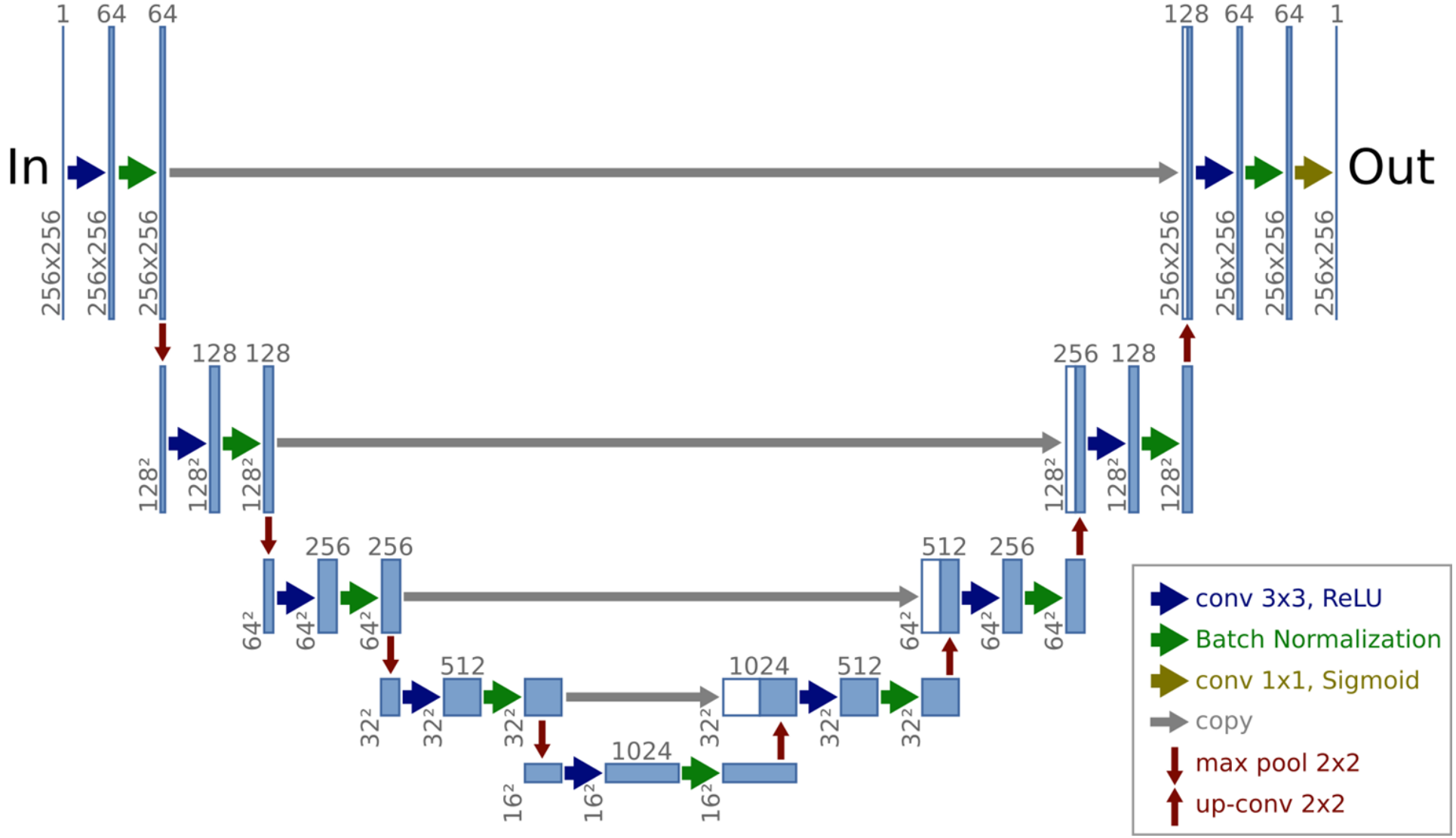
First approach:

Border detection

U-Net



Cardbox detection: using a “U-net”



« All » functionals are by now approximable by a neural network, provided enough examples are given. The only left variational problem is the minimization of the loss functional of the neural network. This loss has two terms: the fidelity term and a regularizing term (typically L2 norm of the coefficients)

Theorem 1 (Universal Approximation Theorem for Width-Bounded ReLU Networks). *For any Lebesgue-integrable function* $f : \mathbb{R}^n \to \mathbb{R}$ *and any* $\epsilon > 0$*, there exists a fully-connected ReLU network* $\mathscr{A}$ *with width* $d_m \leq n + 4$*, such that the function* $F_{\mathscr{A}}$ *represented by this network satisfies*

$$\int_{\mathbb{R}^n} |f(x) - F_{\mathscr{A}}(x)| \mathrm{d}x < \epsilon. \tag{3}$$

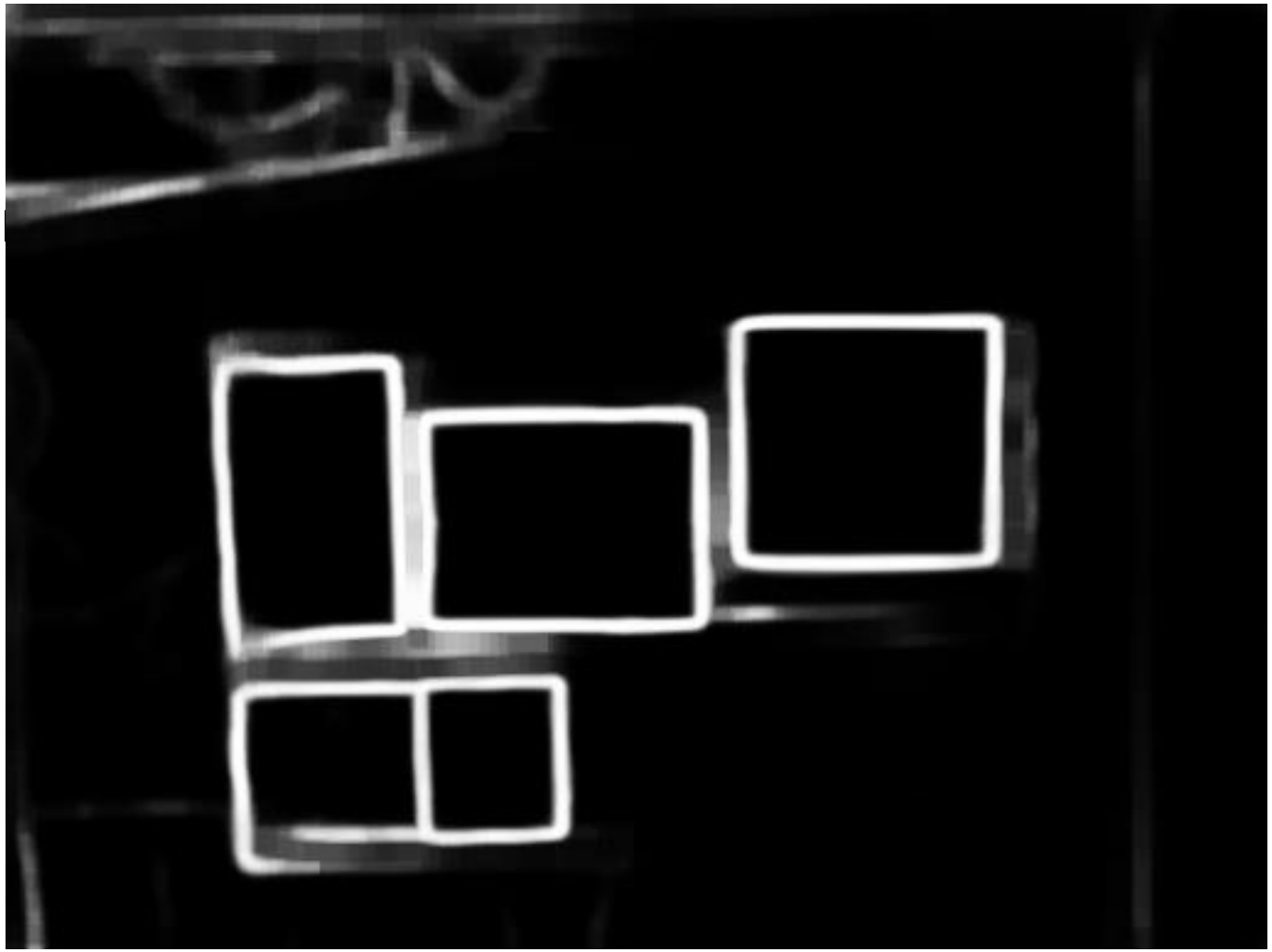
Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The expressive power of neural networks: A view from the width.
Advances in neural information processing systems (pp. 6231-6239).

Cardbox detection

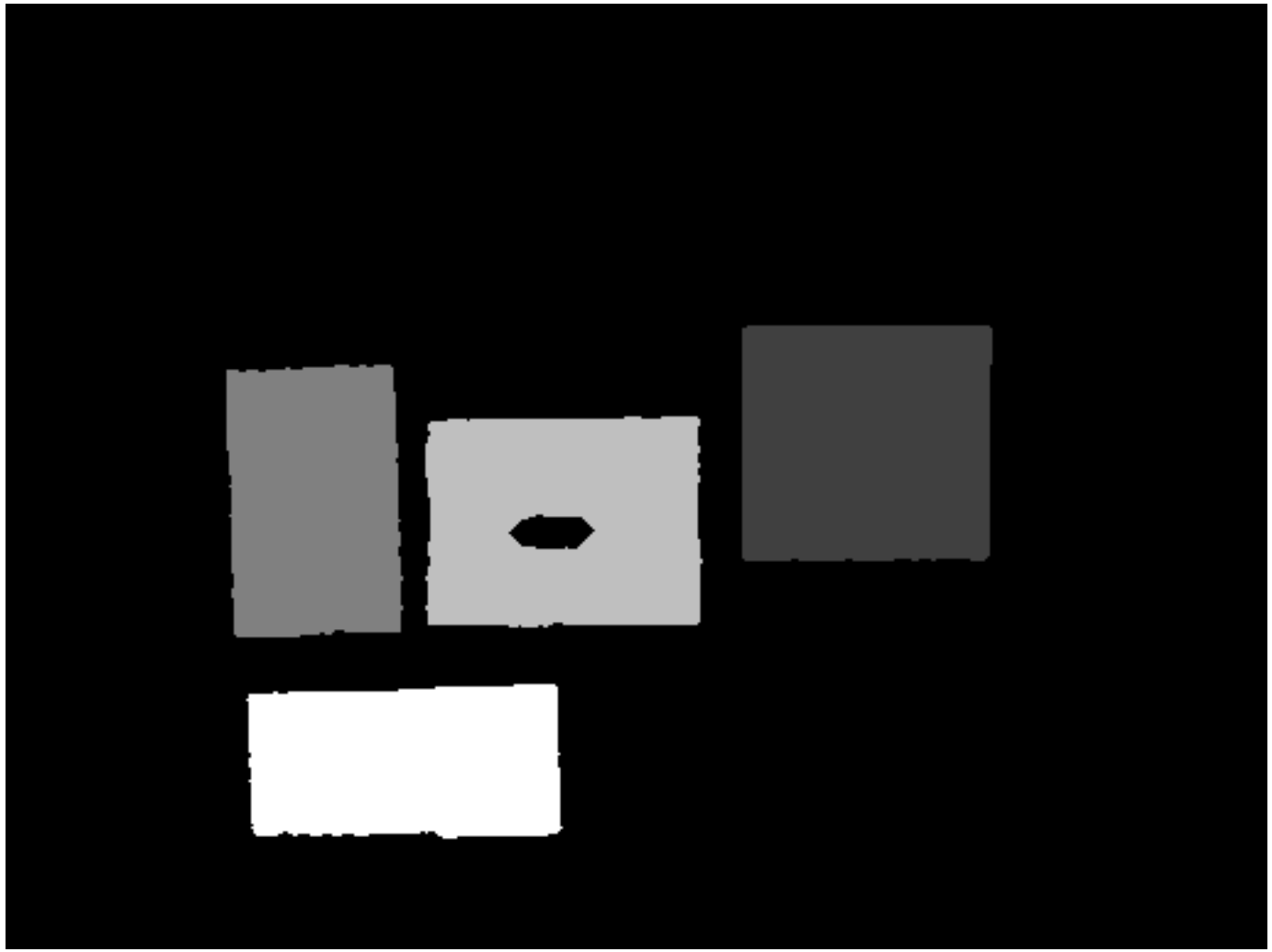
First approach:

Border detectio

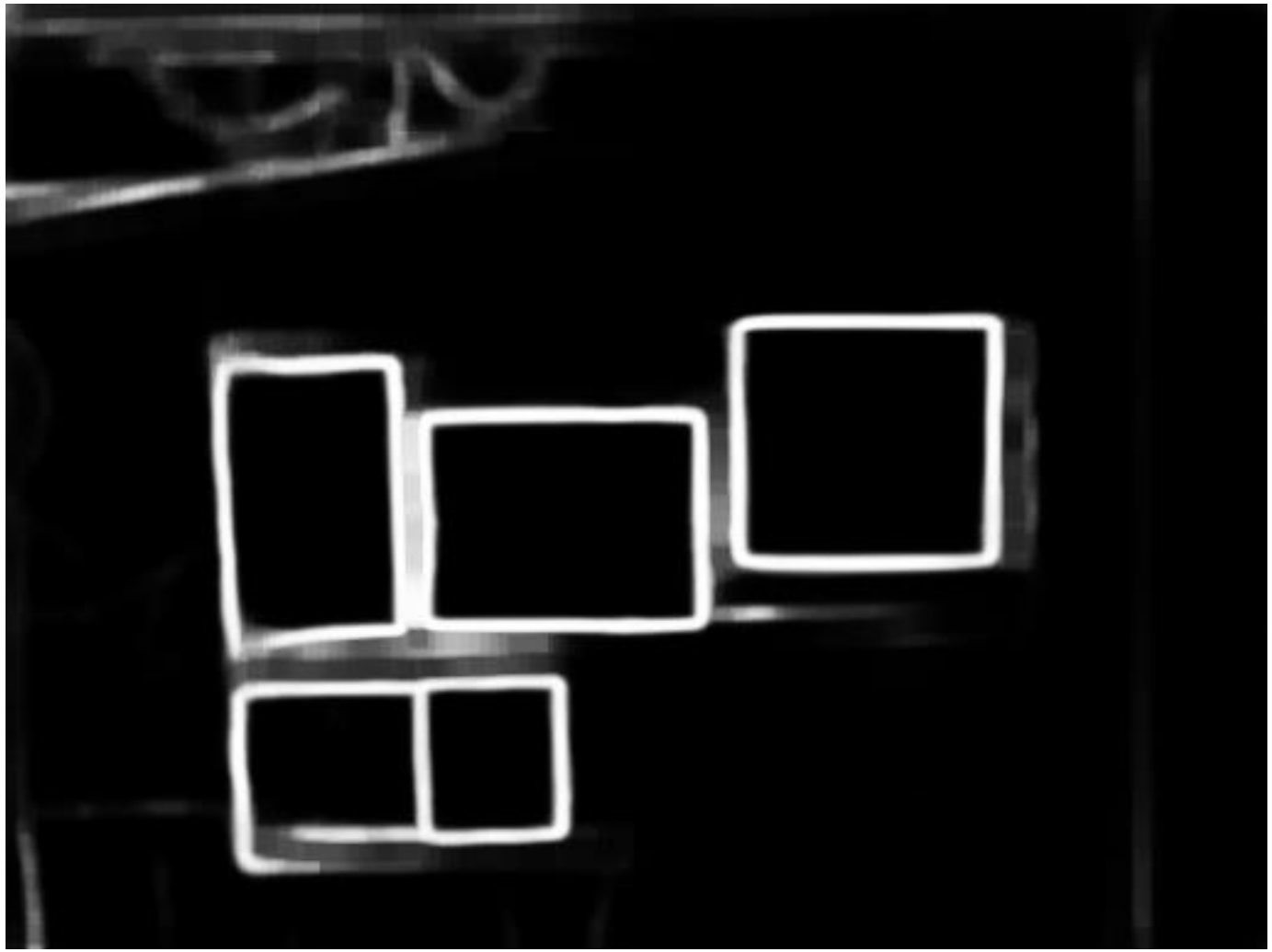
U-Net



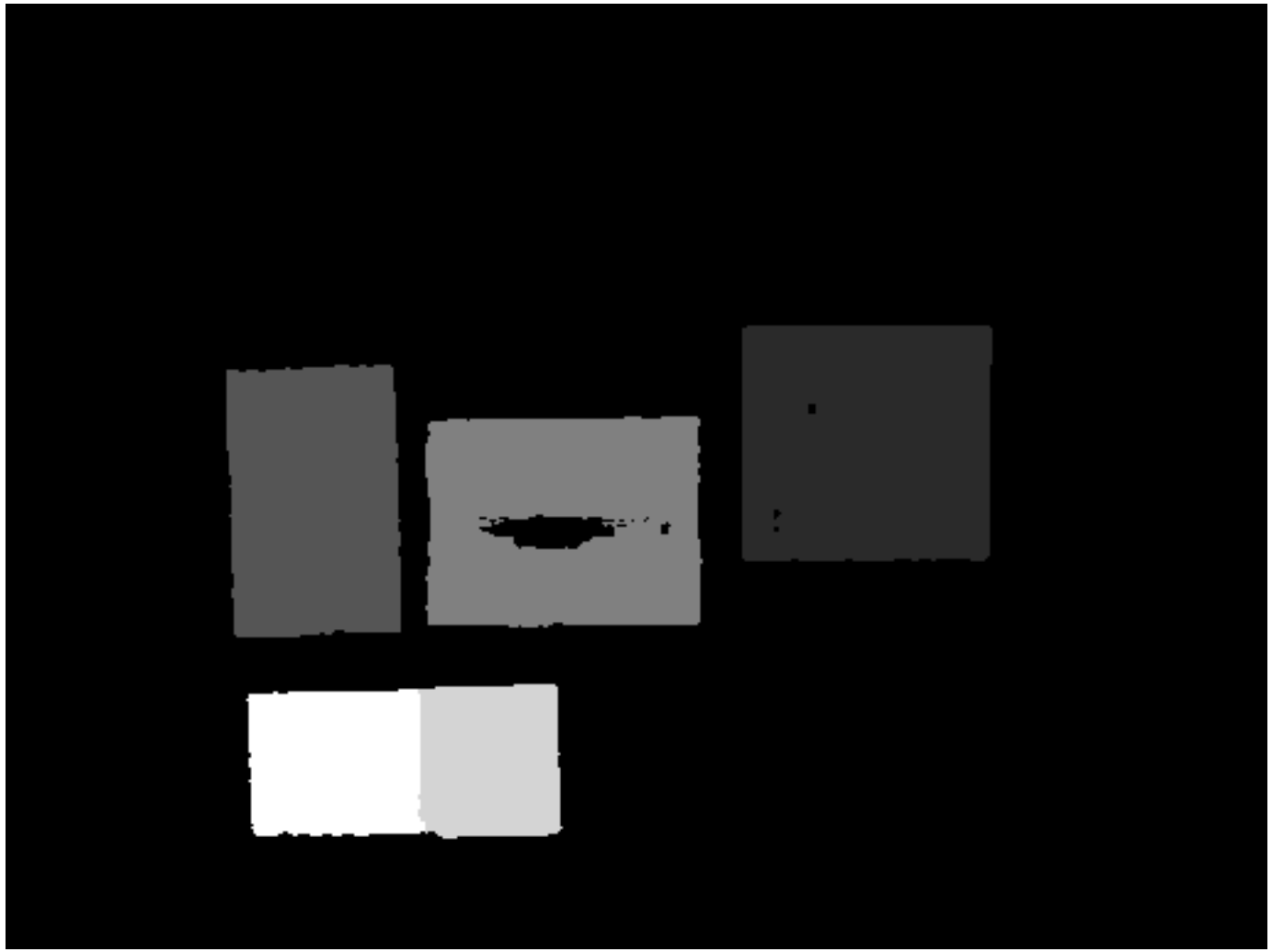
Cardbox detection



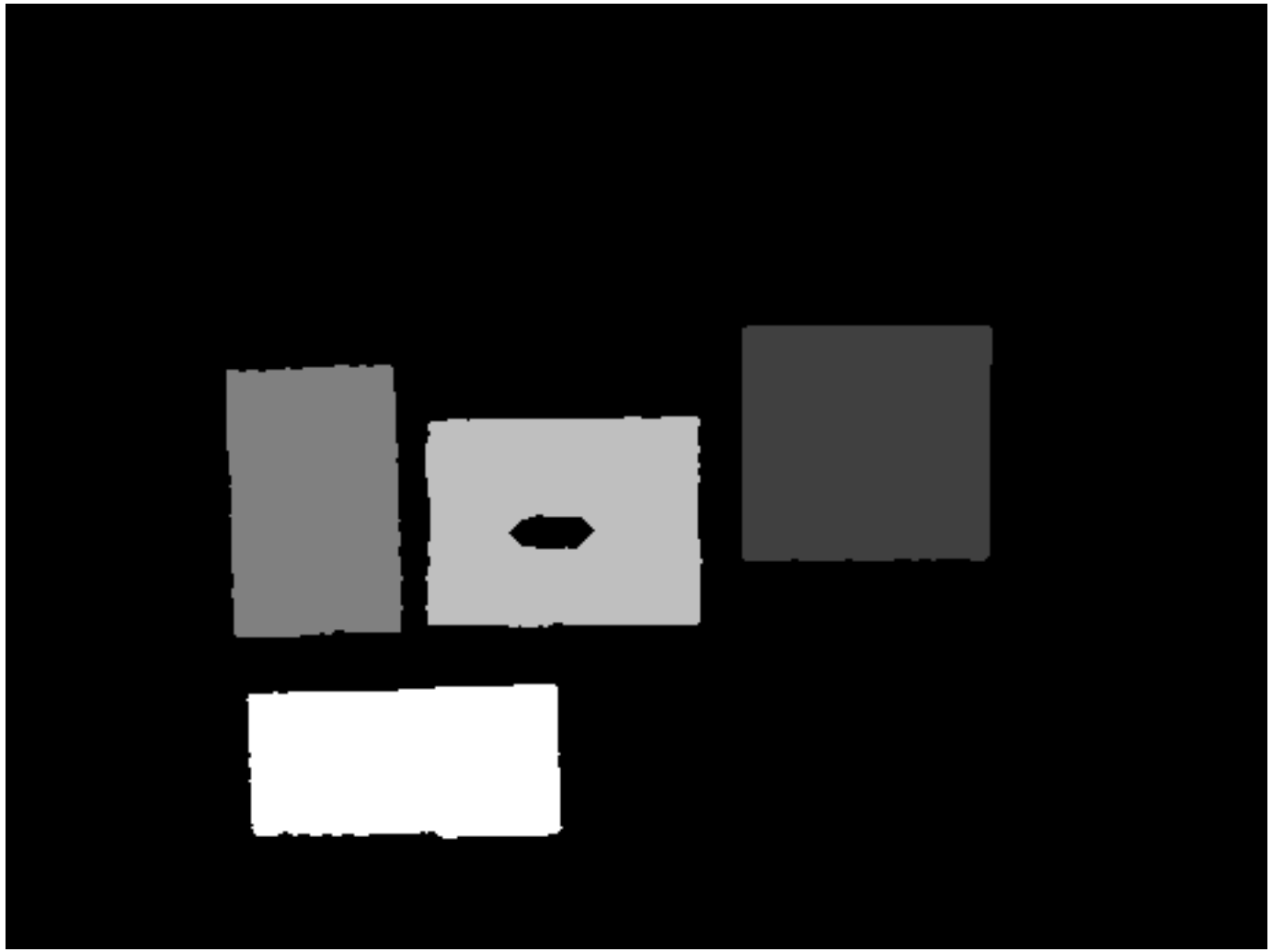
Cardbox detection



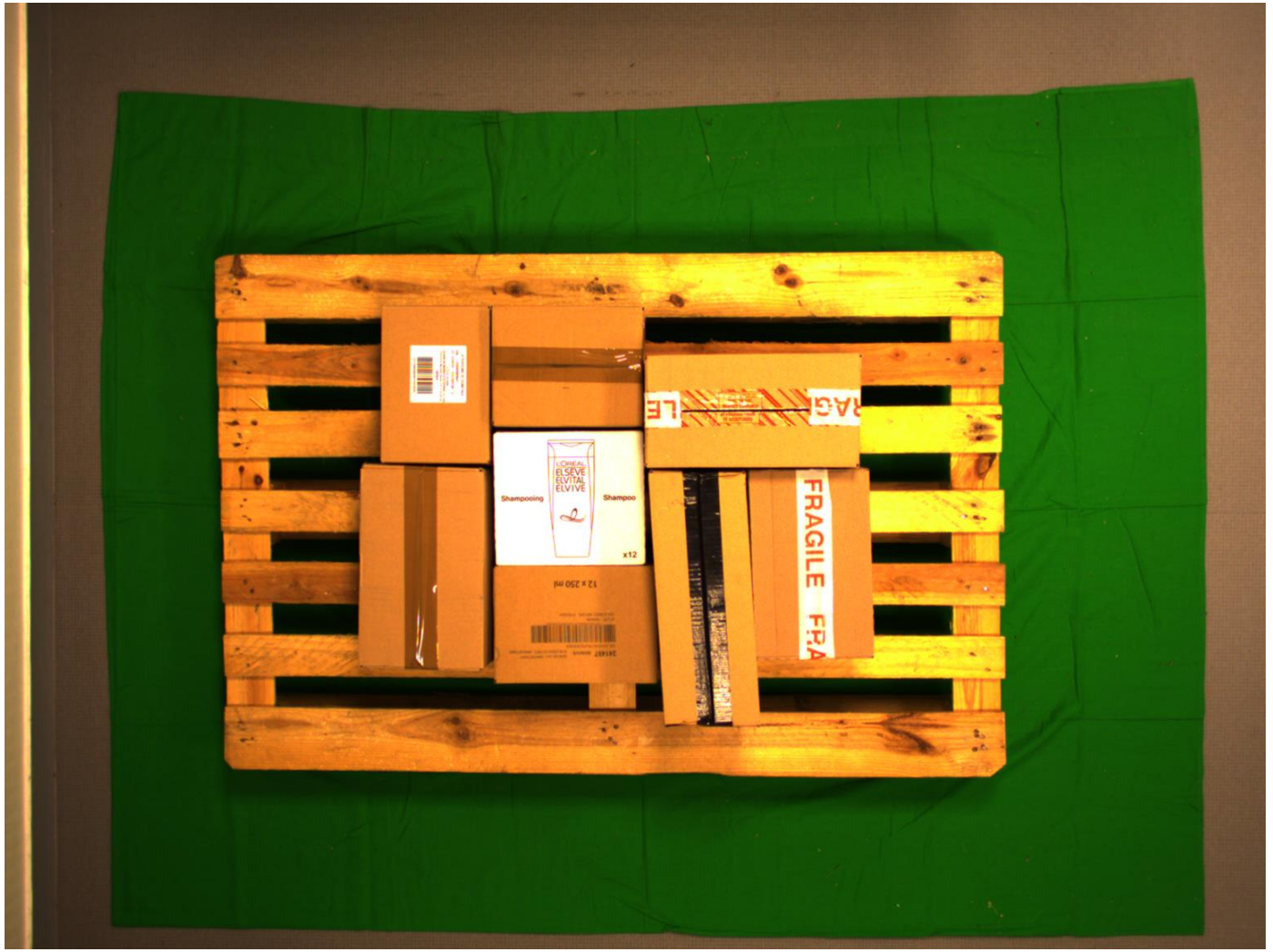
Cardbox detection



Cardbox detection



Cardbox detection



Cardbox detection

→ Detection algorithm should be as robust as possible to changes in scale, rotation, contrast, illumination, shadows…

Making the learning data:

- Acquiring and labeling as many images as possible:
  - ~ 2000 images
  - Limitations: diversity!
- Choosing adequate data augmentation.
- Choosing adequate loss function / neural network structure.
- Adding simulation images for increasing diversity.

Cardbox detection

Unity3D

Texture of cardboxes taken from our existing dataset (+ add changes in the future).

Random texture in background.

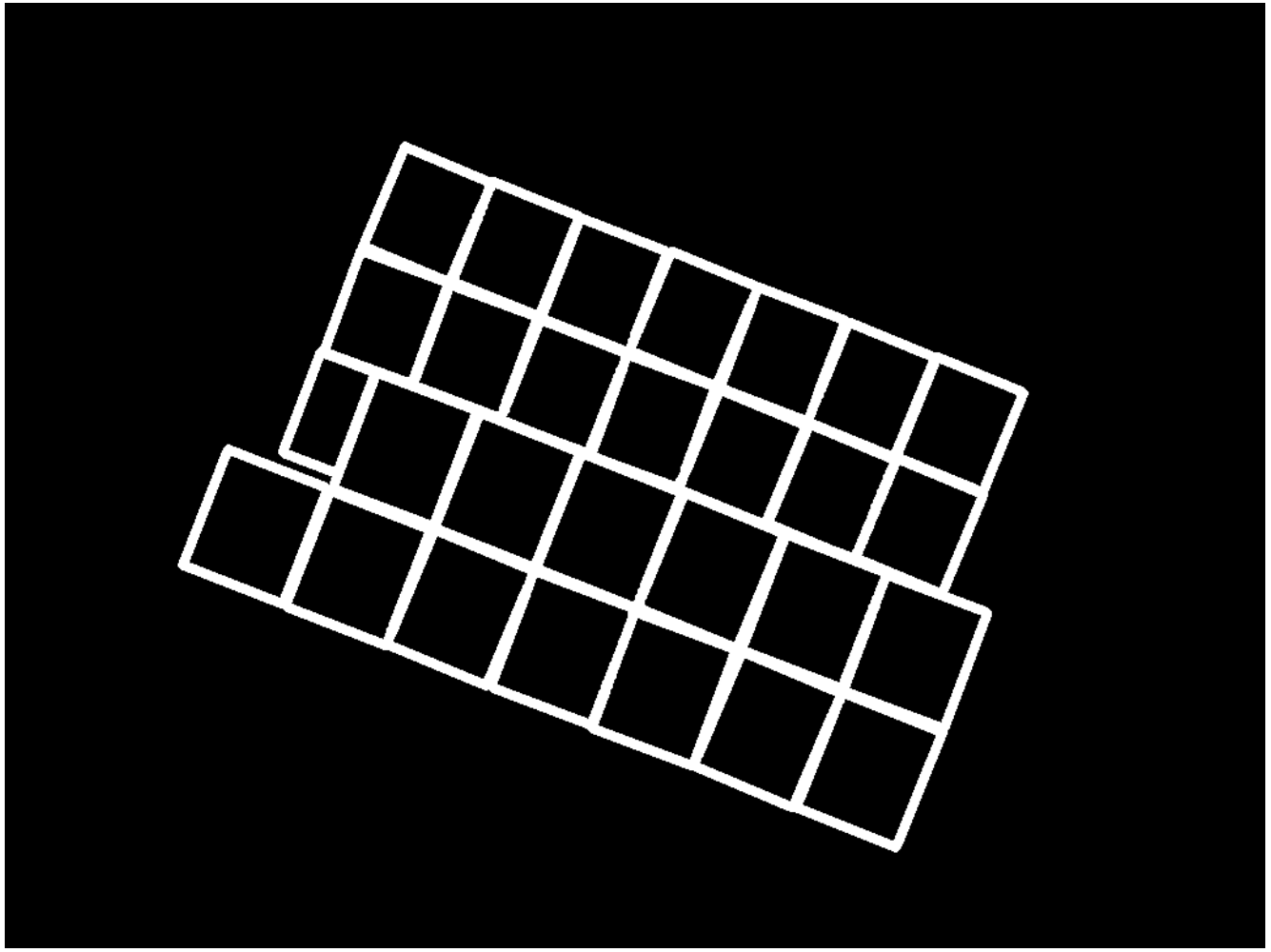
Random orientation, illumination...

Cardbox detection

Unity3D

Cardbox detection

Unity3D

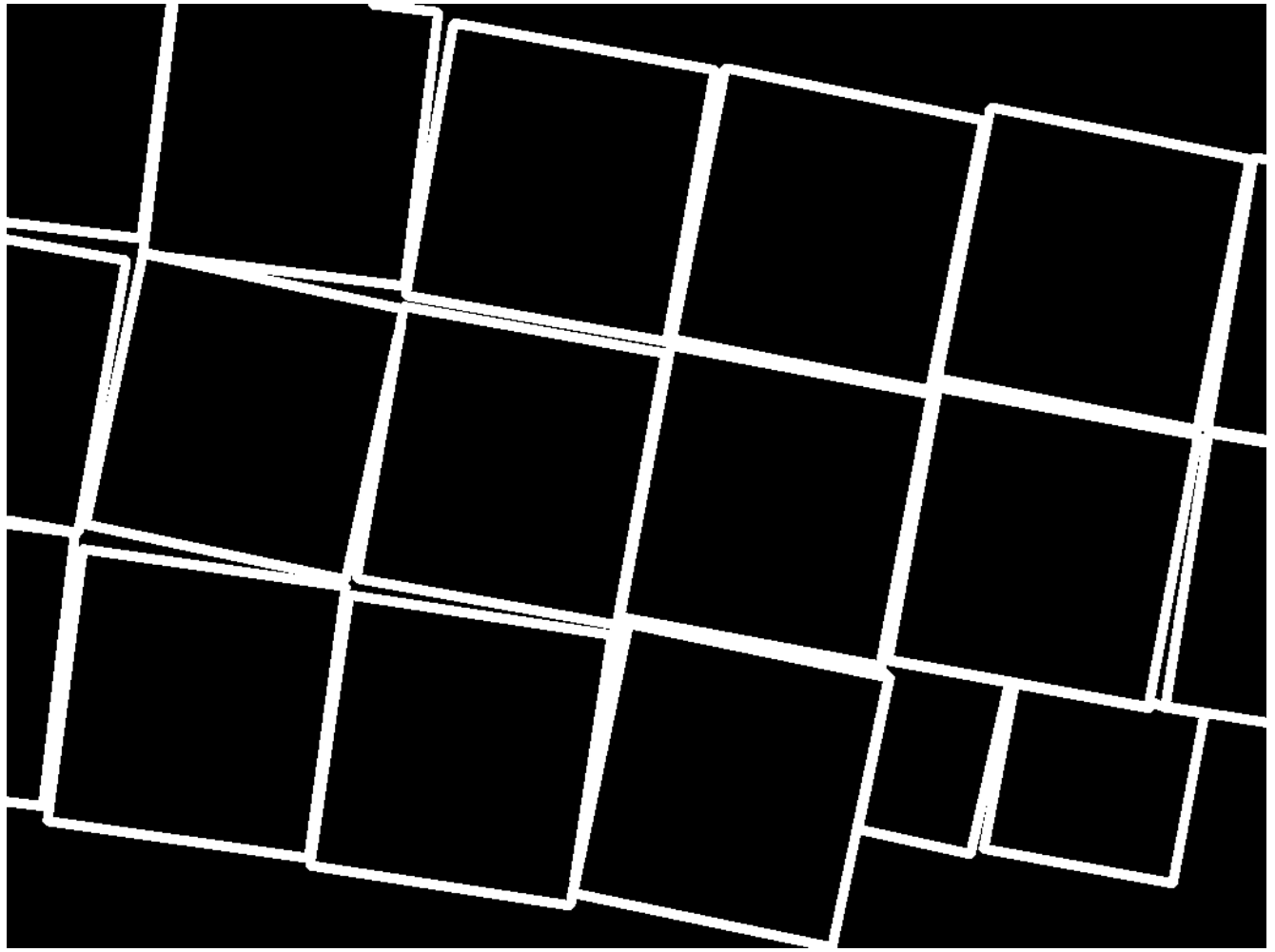


Cardbox detection

Unity3D

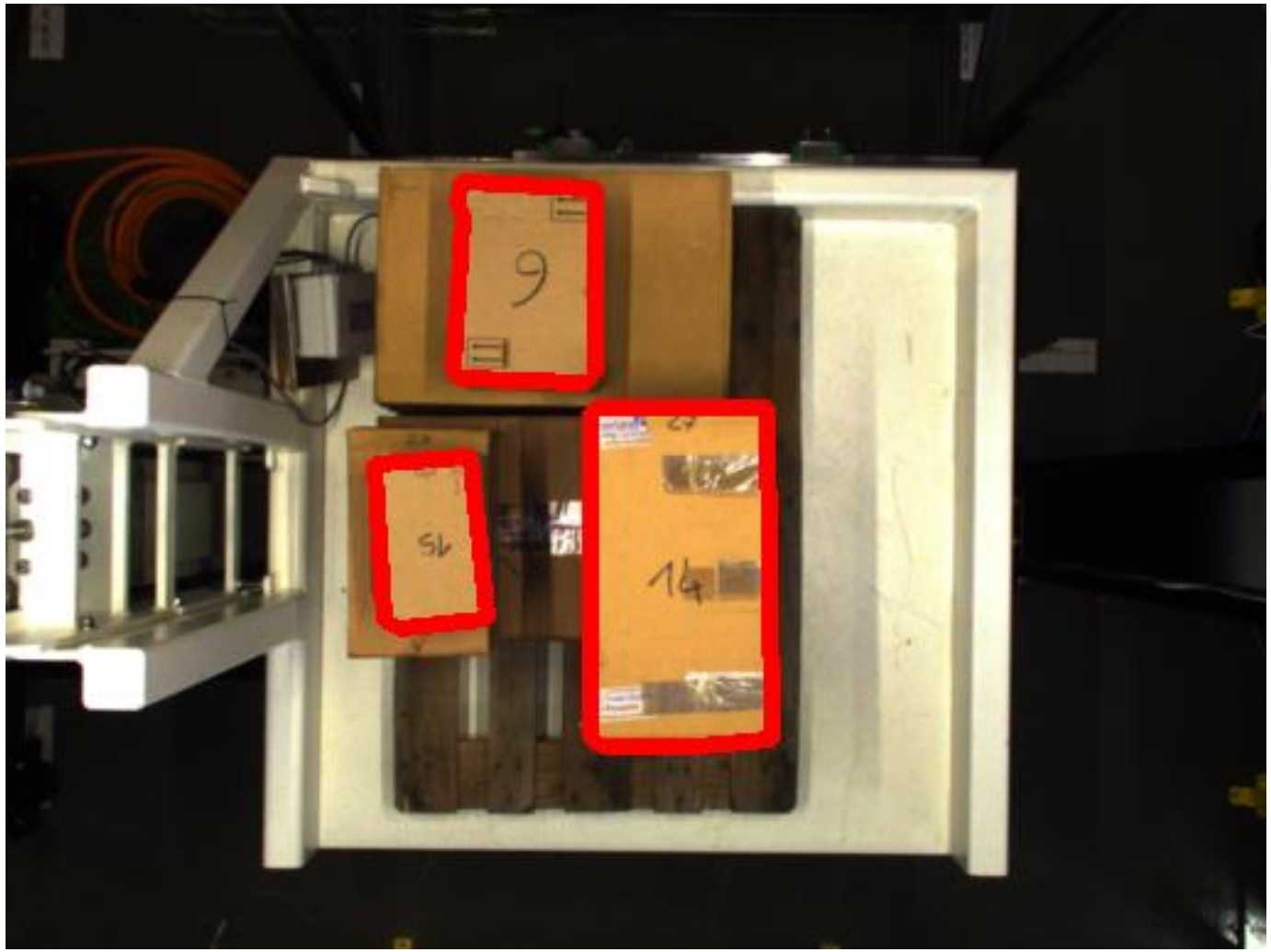
Cardbox detection

Unity3D



Cardbox detection

Results on new environment.

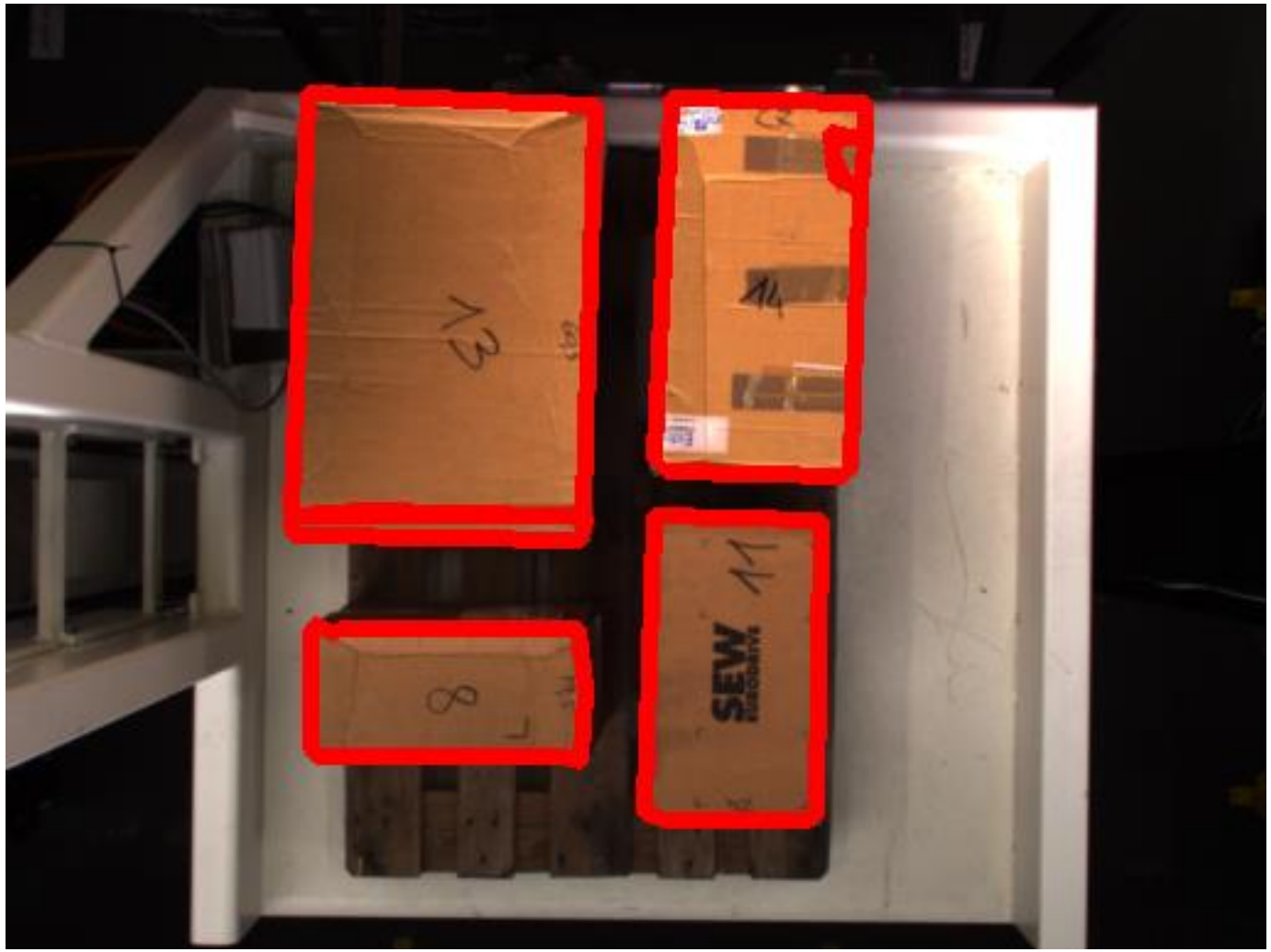


Cardbox detection

Results on new environment.

Cardbox detection

Results on new environment.

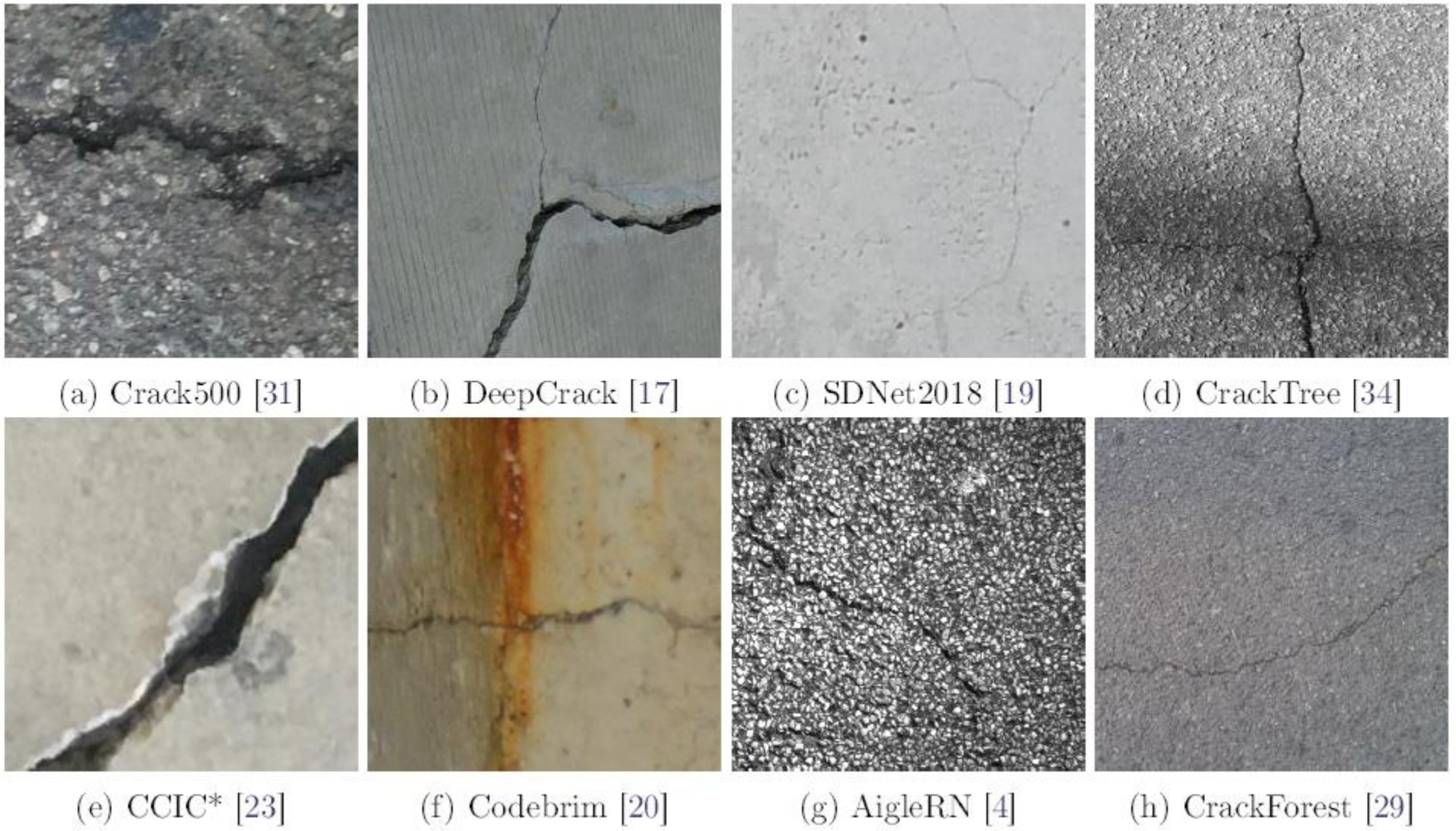


Crack Detection

(Sébastien Drouyer)

An "all terrain" crack detector obtained by deep learning on available databases

Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im



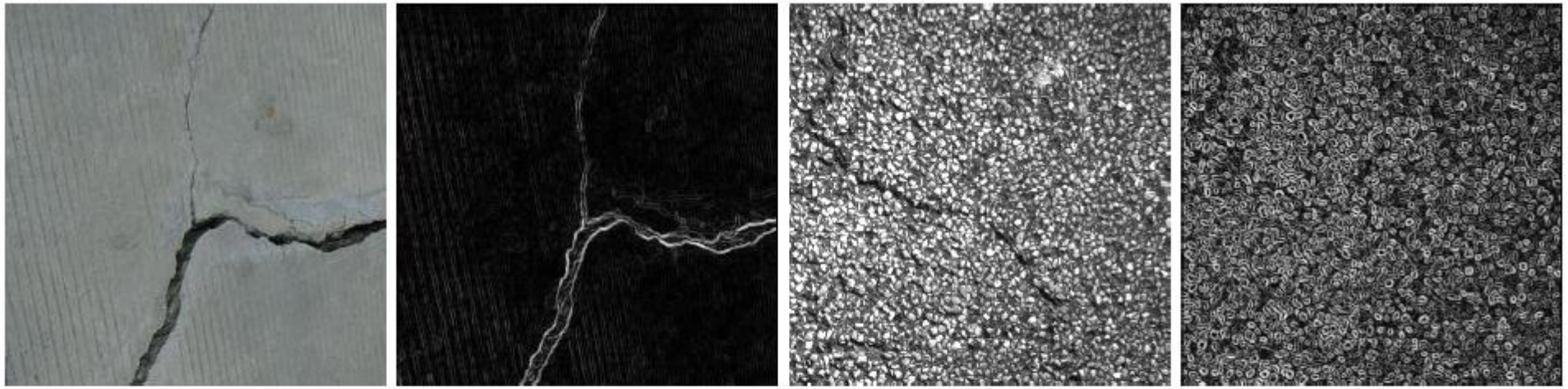
(a) Crack500 [31] (b) DeepCrack [17] (c) SDNet2018 [19] (d) CrackTree [34]

(e) CCIC* [23] (f) Codebrim [20] (g) AigleRN [4] (h) CrackForest [29]

Figure 1: Crops of different crack datasets. CCIC: Concrete Crack Images for Classification.

An "all terrain" crack detector obtained by deep learning on available databases

Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im



(a) Easy example (b) Sobel applied on (a) (c) Hard example (d) Sobel applied on (c)

Figure 2: Results of the Sobel operator on easy and hard examples.

$$E_\infty(\Gamma) = \int_\Gamma \left[v_\infty - \left(\frac{\partial g}{\partial n} \right)^2 \right] ds$$

An "all terrain" crack detector obtained by deep learning on available databases

Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im

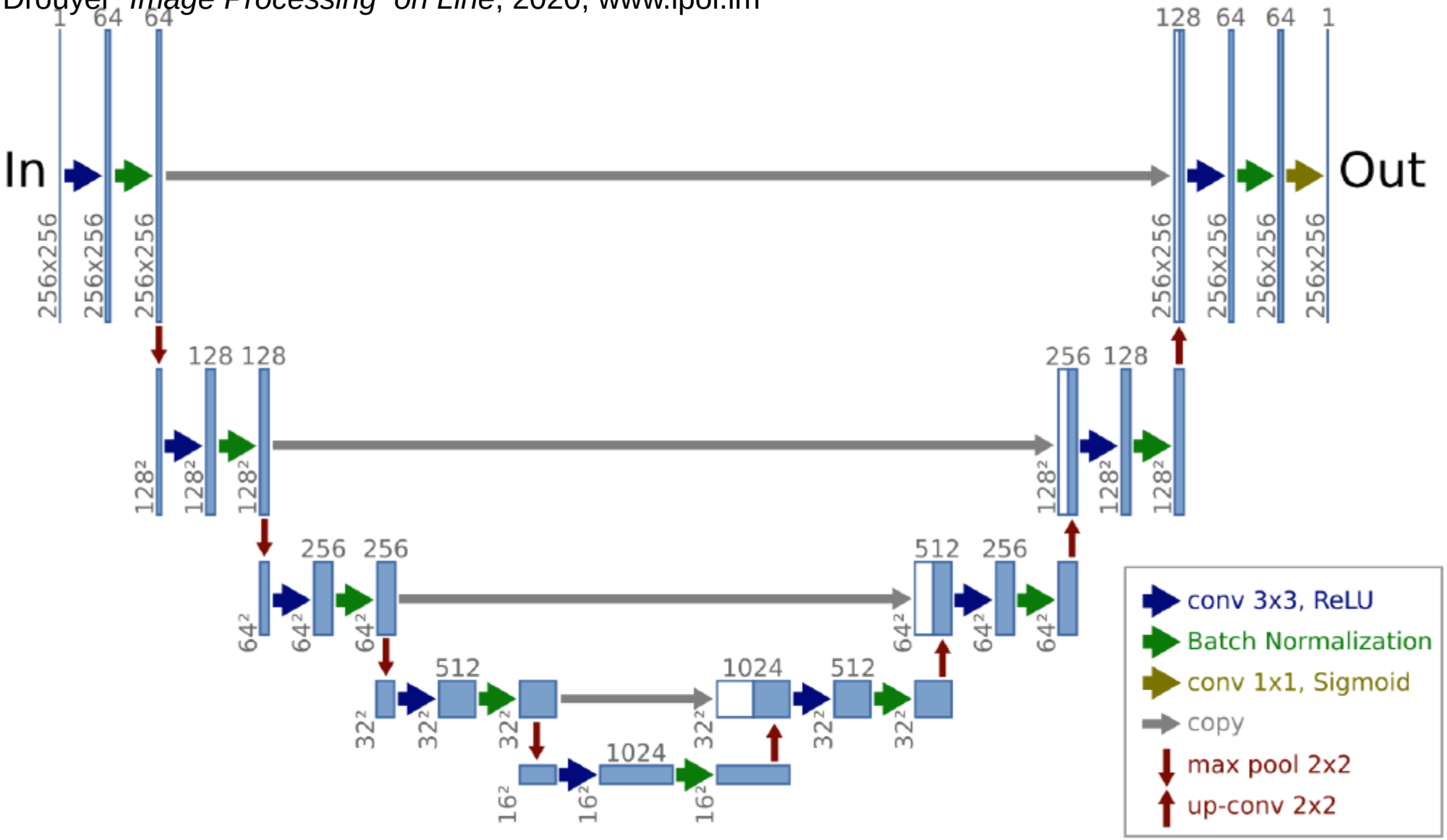


Figure 3: Architecture of our model. In: the input image. Out: the output semantic segmentation.

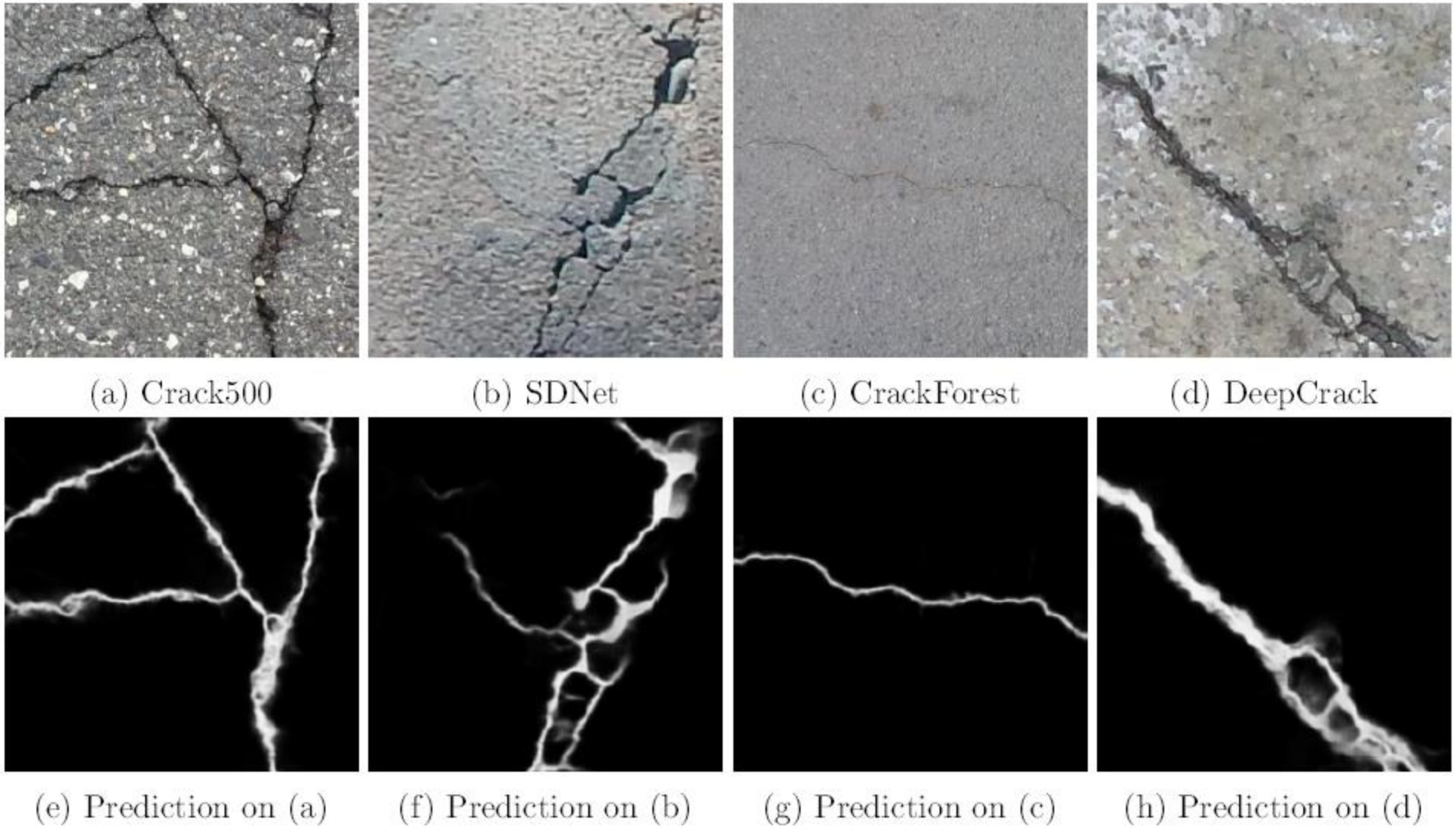
An "all terrain" crack detector obtained by deep learning on available databases
Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im

Method	**Precision**	**Recall**	**F1**	**θ-Precision**	**θ-Recall**	**θ-F1**
CrackIT	*59%*	*26%*	36%	*68%*	*43%*	52%
DeepCrack	*50%*	*80%*	61%	*74%*	*88%*	80%
FCN8	*76%*	*52%*	61%	*88%*	*76%*	82%
FCN32	*73%*	*20%*	31%	*88%*	*31%*	46%
PSPNet	*65%*	*21%*	32%	*84%*	*35%*	49%
SegNet	*72%*	*47%*	57%	*88%*	*63%*	74%
Our method	*77%*	*65%*	70%	*88%*	*86%*	87%

Table 7: Performance metrics for the binary segmentation map for several methods. $\theta = 5$.

An "all terrain" crack detector obtained by deep learning on available databases

Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im



(a) Crack500 (b) SDNet (c) CrackForest (d) DeepCrack

(e) Prediction on (a) (f) Prediction on (b) (g) Prediction on (c) (h) Prediction on (d)

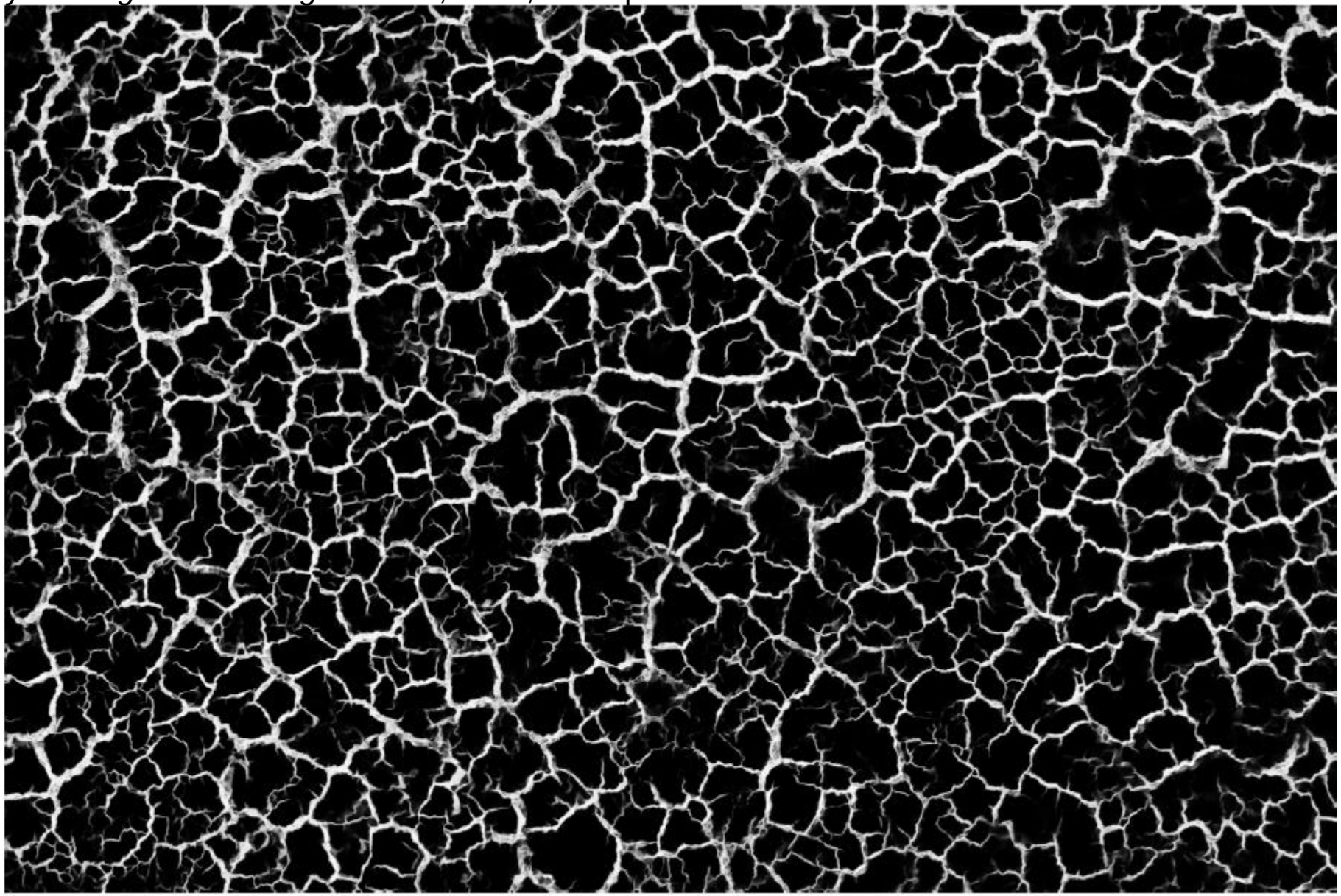
Figure 7: Examples of predictions of the model trained on the merged dataset.

An "all terrain" crack detector obtained by deep learning on available databases

Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im

An "all terrain" crack detector obtained by deep learning on available databases

Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im



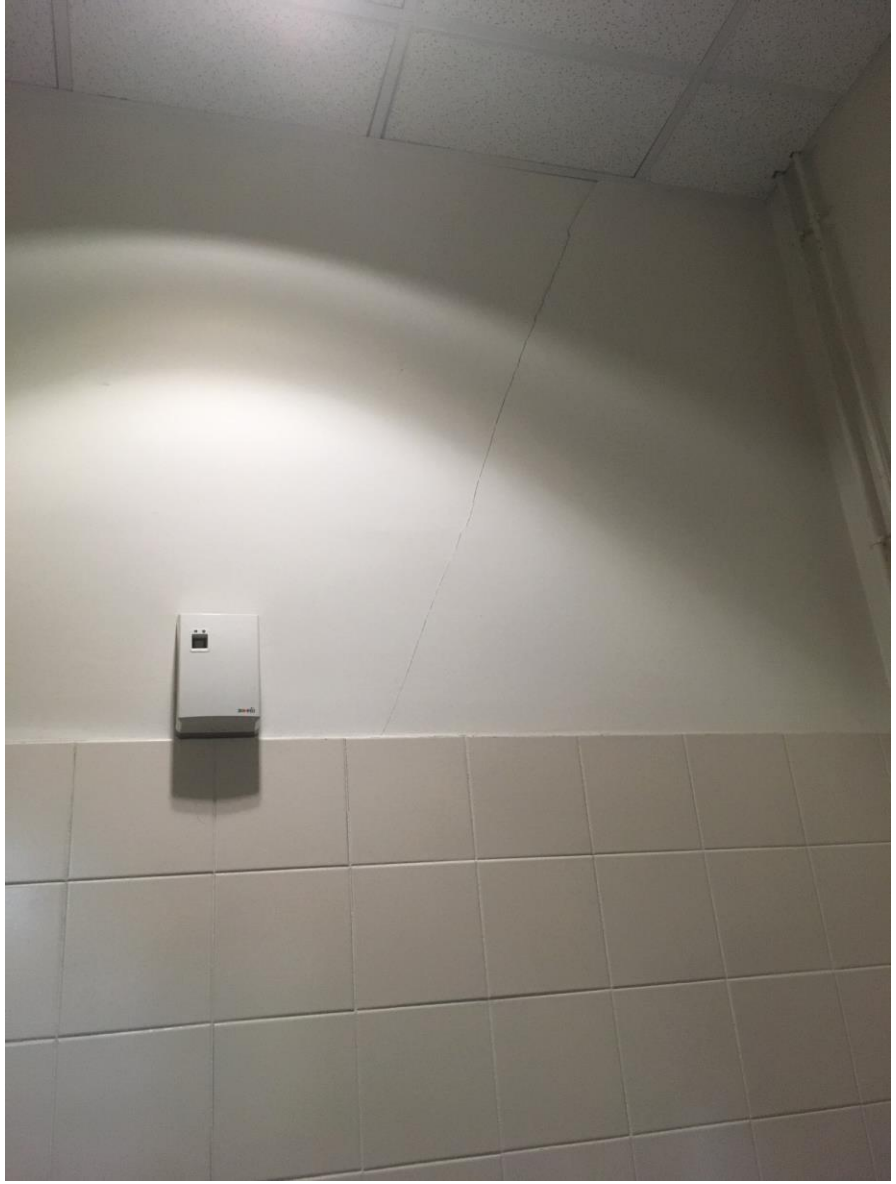
An "all terrain" crack detector obtained by deep learning on available databases
Sébastien Drouyer *Image Processing on Line*, 2020, www.ipol.im

(a) Image of a spider

(b) Detection when trained on original dataset

(c) Detection when trained on augmented dataset

Figure 10: Example of problematic image: an image of a spider on a wall is detected as a crack when using U-net trained on the original merged dataset. However, when using the augmented dataset, our detector correctly classify the spider as non crack.



S. Drouyer An 'All Terrain' Crack Detector Obtained by Deep Learning on Available Databases IPOL 2019 http://www.ipol.im/pub/pre/282/

S. Drouyer An 'All Terrain' Crack Detector Obtained by Deep Learning on Available Databases IPOL 2019 http://www.ipol.im/pub/pre/282/

Discussion: Bayesian framework? Calculus of Variations?

Data:

$$(x_i, y_i)_{i \in I}$$

Functional:

$$y = F(\theta, x)$$

Minimization:

$$\arg\min_{\theta} \sum_i ||y_i - F(\theta, x_i)||^2 + ||\theta||^2$$

(Data fit + regularity term that actually fixes the regularity of the functional itself).

A standard feed-forward network architecture is used to implement FFDNet, as shown in Figure 1. The network is composed of D convolutional layers, which share the same structure. Each of these has W outputs, and the spatial size of their kernels is $K \times K$.

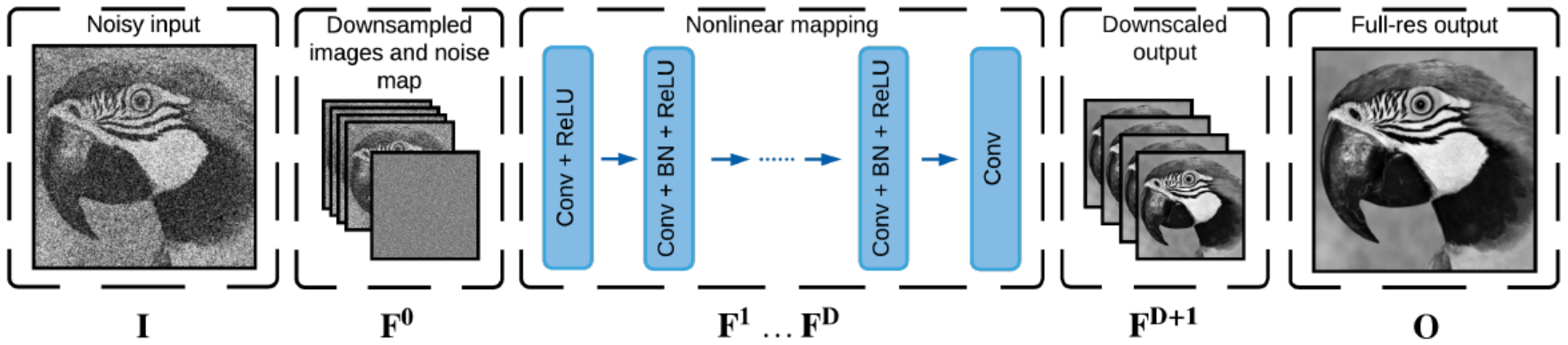
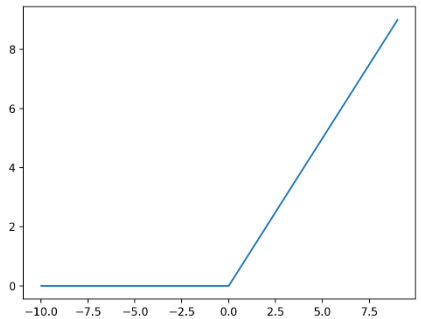


Figure 1: Architecture of FFDNet.



Discussion

The « general » functional generally also has a Bayesian interpretation

Bayesian patch-based methods

8.1 The principle

Given u the noiseless ideal image and $\tilde{u}$ the noisy image corrupted with Gaussian noise of standard deviation σ so that

$$\tilde{u} = u + n, \tag{8.1}$$

the conditional distribution $\mathbb{P}(\tilde{u} \mid u)$ is

$$\mathbb{P}(\tilde{u} \mid u) = \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} e^{-\frac{||u-\tilde{u}||^2}{2\sigma^2}}, \tag{8.2}$$

where M is the total number of pixels in the image.

Discussion

The « general » functional often has a Bayesian interpretation

Example: denoising

Global image denoising

$$\mathbb{P}(\tilde{P} \mid P) = \frac{1}{(2\pi\sigma^2)^{\frac{\kappa^2}{2}}} e^{-\frac{||P-\tilde{P}||^2}{2\sigma^2}}, \tag{8.11}$$

Bayesian minimum squared error (MMSE) is by Bayes' formula

$$\hat{P} = \mathbb{E}[P \mid \tilde{P}] = \int \mathbb{P}(P \mid \tilde{P}) P dP = \int \frac{\mathbb{P}(\tilde{P} \mid P)}{\mathbb{P}(\tilde{P})} \mathbb{P}(P) P dP. \tag{8.12}$$

Using a huge set of M natural patches (with a distribution supposedly approximating the real natural patch density), we can approximate the terms in (8.12) by $\mathbb{P}(P)dP \simeq \frac{1}{M}$ and $\mathbb{P}(\tilde{P}) \simeq \frac{1}{M} \sum_i \mathbb{P}(\tilde{P} \mid P_i)$, which in view of (8.11) yields

$$\hat{P} \simeq \frac{\frac{1}{M} \sum_i \mathbb{P}(\tilde{P} \mid P_i) P_i}{\frac{1}{M} \sum_i \mathbb{P}(\tilde{P} \mid P_i)}.$$

The Levin and Nadler optimal « global denoising algorithm » uses « all patches of the world »

- **Input**: Noisy image $\tilde{u}$, its patches $\tilde{P}$
- **Input**: Very large set of $M = 2^{10}$ patches P_i extracted from a large set of noiseless natural images (20000)
- **Output**: Denoised image $\hat{u}$.
- **for all patches $\tilde{P}$ extracted from $\tilde{u}$:** Compute the MMSE denoised estimate of $\tilde{P}$

$$\hat{P} \simeq \frac{\sum_{i=1}^{M} \mathbb{P}(\tilde{P} \mid P_i) P_i}{\sum_{i=1}^{M} \mathbb{P}(\tilde{P} \mid P_i)} \qquad G_\sigma^N(\mathbf{x}) = \frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{-\frac{||\mathbf{x}||^2}{2\sigma^2}}$$

  where $\mathbb{P}(\tilde{P} \mid P_i)$ is known from the Gaussian noise distribution.
- (Aggregation) : for each pixel $\mathbf{j}$ of u, compute the denoised version $\hat{u}_\mathbf{j}$ as the average of all values $\hat{P}(\mathbf{j})$ for all patches

A. Levin, B. Nadler. *CVPR* 2011. Natural image denoising: Optimality and inherent bounds
Zoran, D., & Weiss, Y. *ICCV* 2011. From learning models of natural image patches to whole image restoration.

An Analysis and Implementation of the FFDNet Image Denoising Method

Matias Tassano, Julie Delon, Thomas Veit

(a) Clean image

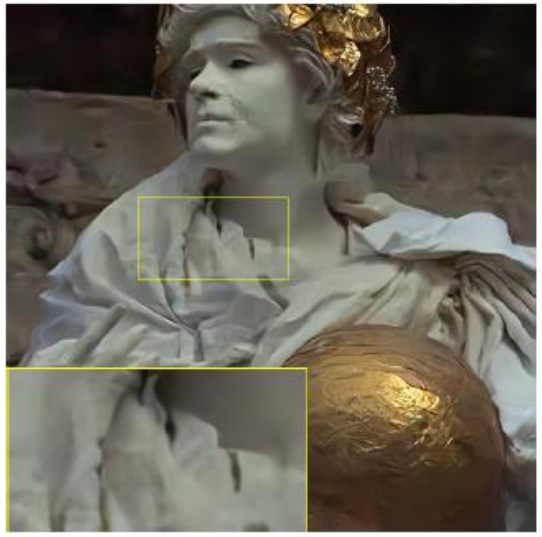
(b) Noisy image, $\sigma_{noise} = 55$

(e) NLB, PSNR 28.26dB

(f) NLDD, PSNR 28.56dB

(c) DnCNN, PSNR 28.82dB

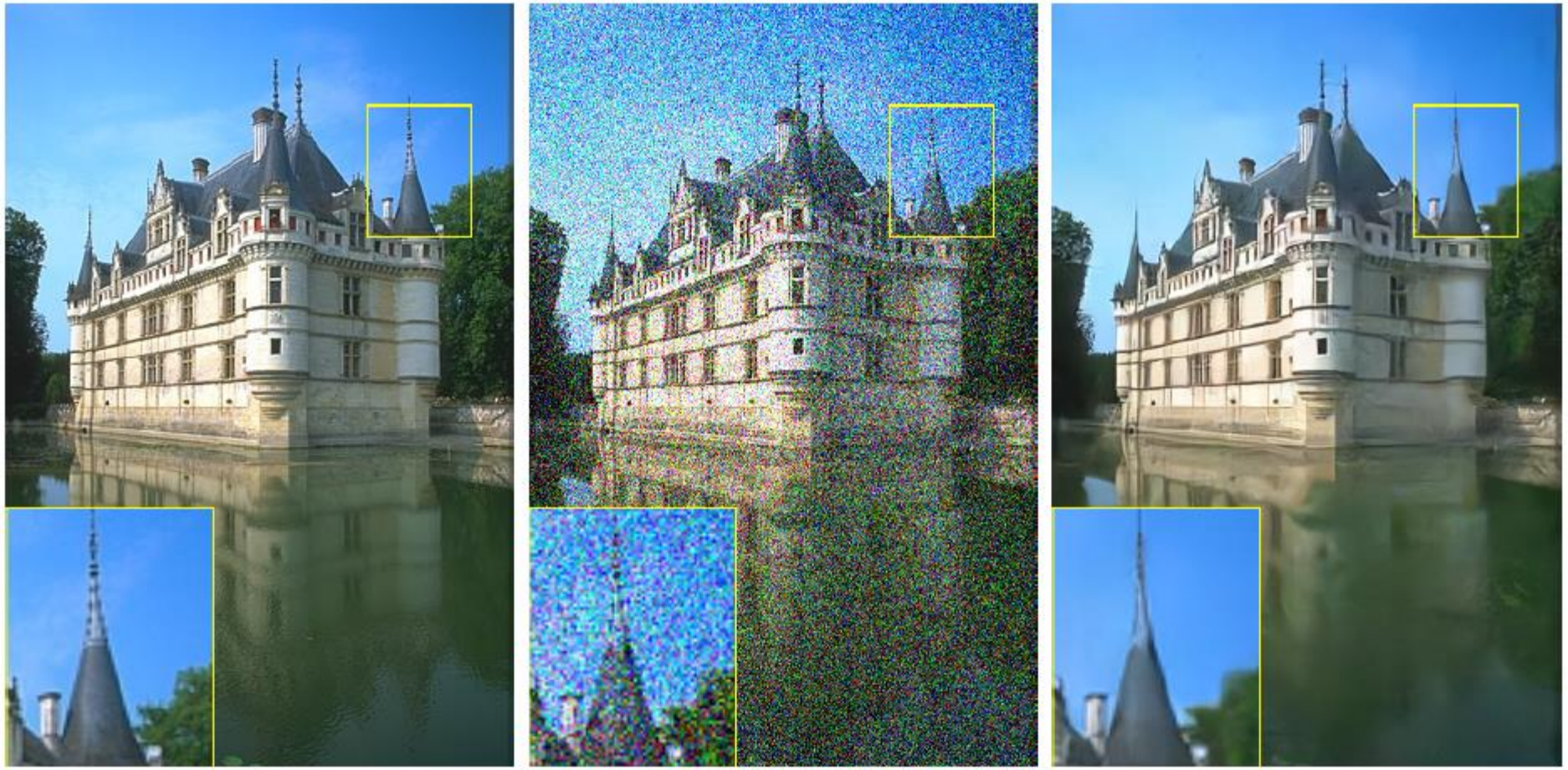
(d) BM3D, PSNR 28.21dB



(g) FFDNet, PSNR 29.08dB

An Analysis and Implementation of the FFDNet Image Denoising Method

Matias Tassano, Julie Delon, Thomas Veit



(a) Clean image (b) Noisy image, $\sigma_{noise} = 55$ (c) FFDNet, PSNR 28.52dB

The training dataset is composed of pairs of input-output patches $\left\{\left((\tilde{\mathbf{I}}_j, \mathbf{M}_j),\, \mathbf{I}_j\right)\right\}_{j=0}^{m}$ which are generated by adding AWG of $\sigma \in [0,\, 75]$ to the clean patches $\mathbf{I}_j$ and building the corresponding noise map $\mathbf{M}_j$ (which is in this case constant with all its elements equal to σ). A total of $m = 128 \times 8000$ patches are extracted from the Waterloo Exploration Database [14], where the mini-batch size is 128. The patch size is 64×64 and 50×50 for grayscale and color images, respectively. Patches are randomly cropped from randomly sampled images of the training dataset. Data is augmented five times by introducing rescaling by different scale factors and random flips[2]. In the cases in which residual learning is used, the network outputs an estimation of the input noise

$$\mathcal{F}(\,\tilde{\mathbf{I}}\,) = \hat{\mathbf{N}}\,. \tag{4}$$

Then, the denoised image is computed by subtracting the output noise to the noisy input

$$\hat{\mathbf{I}} = \tilde{\mathbf{I}} - \mathcal{F}(\,\tilde{\mathbf{I}}\,)\,. \tag{5}$$

In this case, the loss function is the following

$$\mathcal{L}_{res}(\theta) = \frac{1}{2m} \sum_{j=1}^{m} \left\| \mathcal{F}(\,(\tilde{\mathbf{I}}_j, \mathbf{M}_j);\, \theta\,) - \mathbf{N}_j \right\|^2\,, \tag{6}$$

where θ is the collection of all learnable parameters.

A standard feed-forward network architecture is used to implement FFDNet, as shown in Figure 1. The network is composed of D convolutional layers, which share the same structure. Each of these has W outputs, and the spatial size of their kernels is $K \times K$.

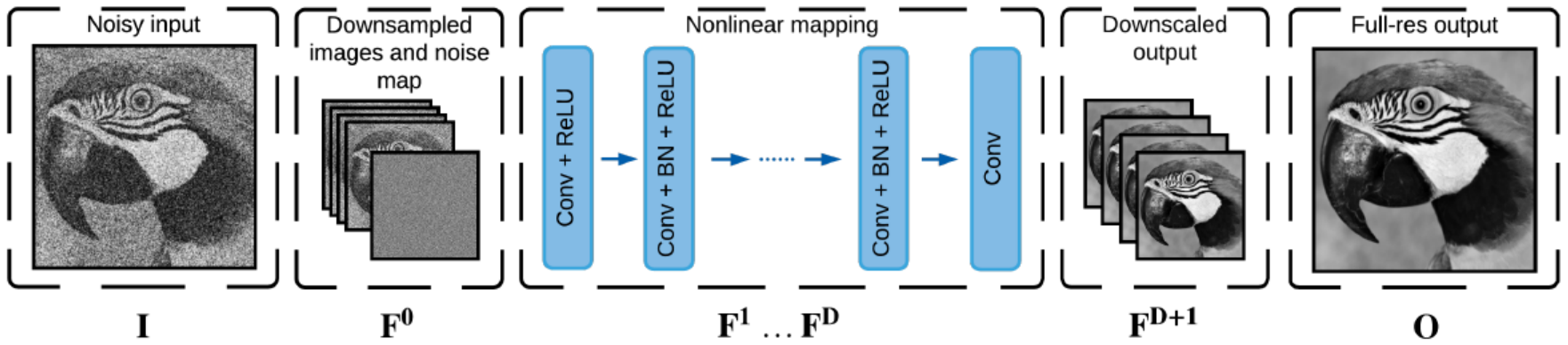
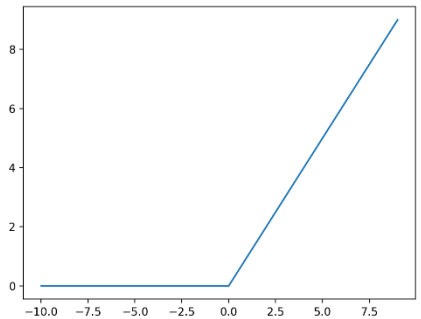


Figure 1: Architecture of FFDNet.



(d) DnCNN, PSNR 28.32dB (e) BM3D, PSNR 27.31dB (f) NLB, PSNR 27.79dB (g) NLDD, PSNR 27.92dB

Comparison of color denoising results