Deep Learning and Our Applications

Sandra Avila

Universidade Federal de Sergipe, 1984

UFS
Universidade Federal de Sergipe

dcomp
departamento de computação

UFMG

Núcleo de Processamento Digital de Imagens / NPDI

UPMC
SORBONNE UNIVERSITÉS

LIP6

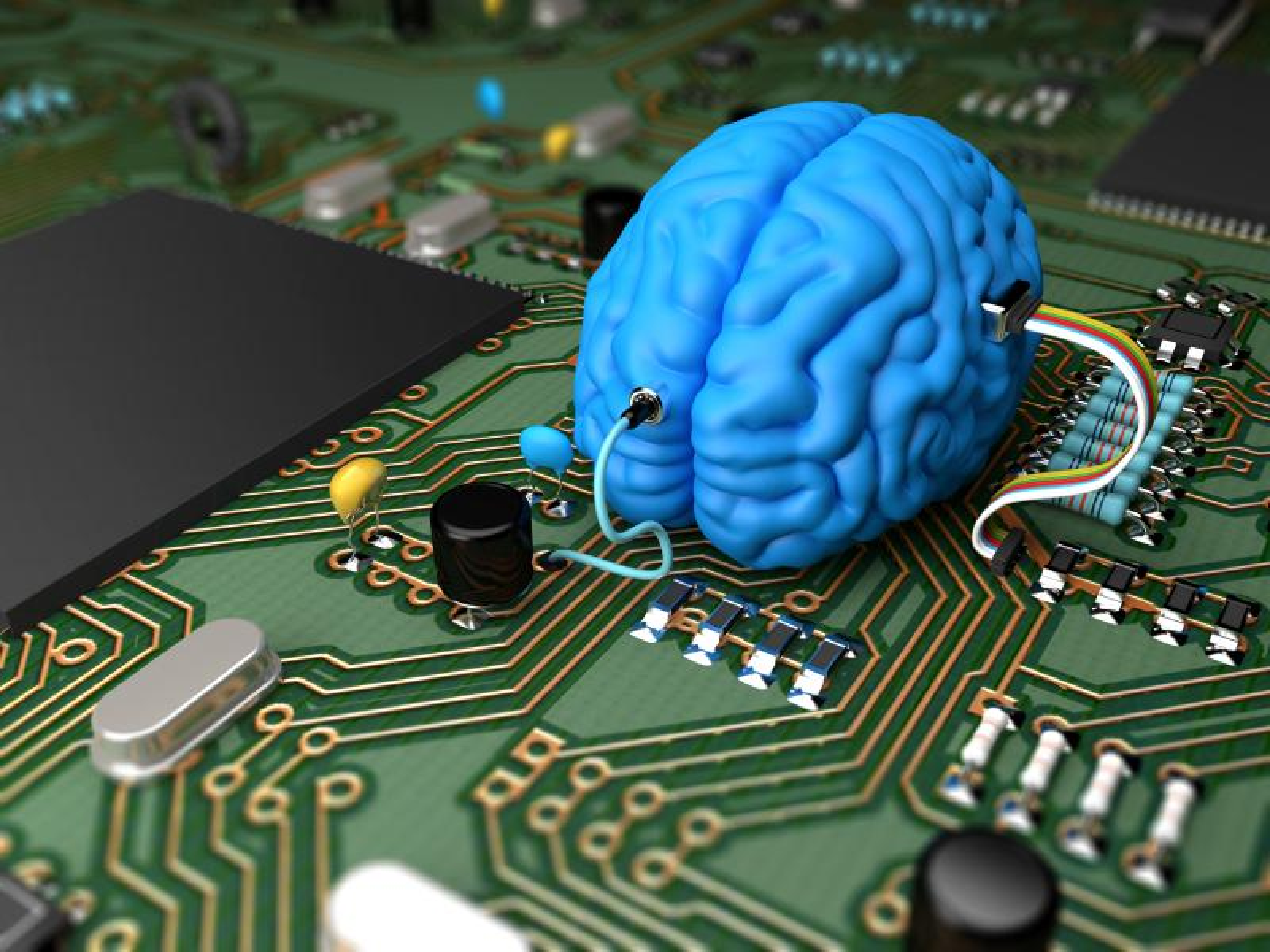
UNICAMP

FEEC

SAMSUNG

RECOD

IEEE Women in Engineering
WIE

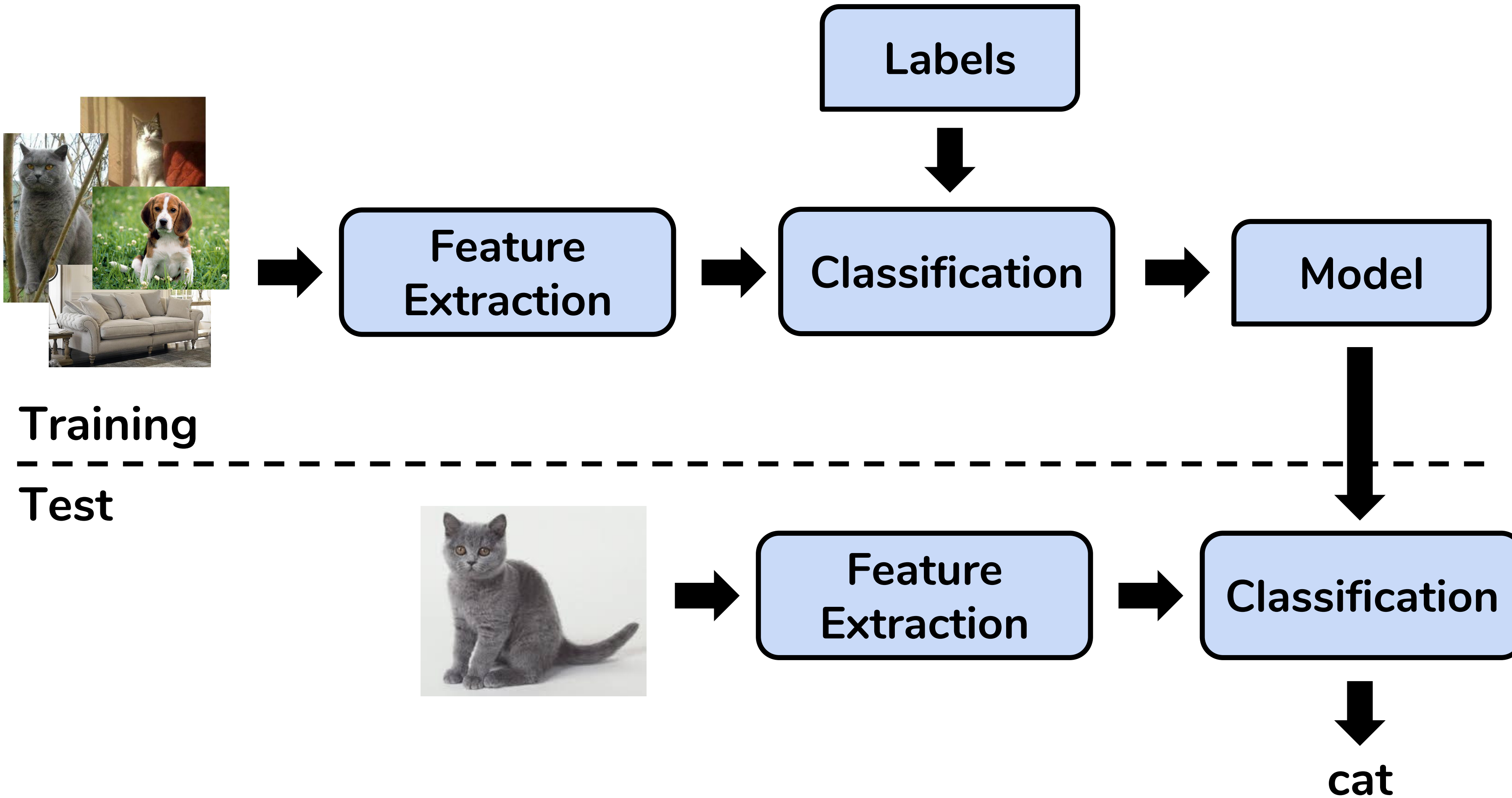


Agenda

- Foundations
- Our Applications
  - Sensitive Media Analysis
  - Melanoma Screening
- Final Considerations

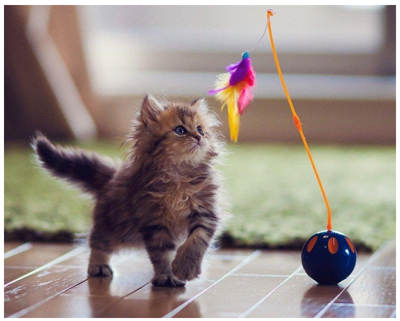
Foundations

Image Classification



Low-level Feature Extraction

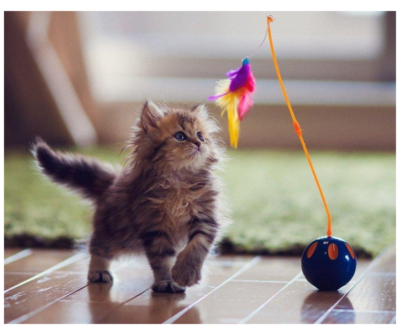
Global Descriptor

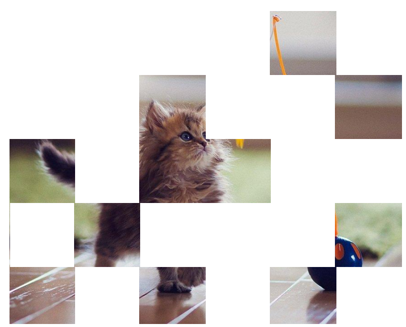


Global Feature Extraction

$$\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{bmatrix}$$

Local Descritor





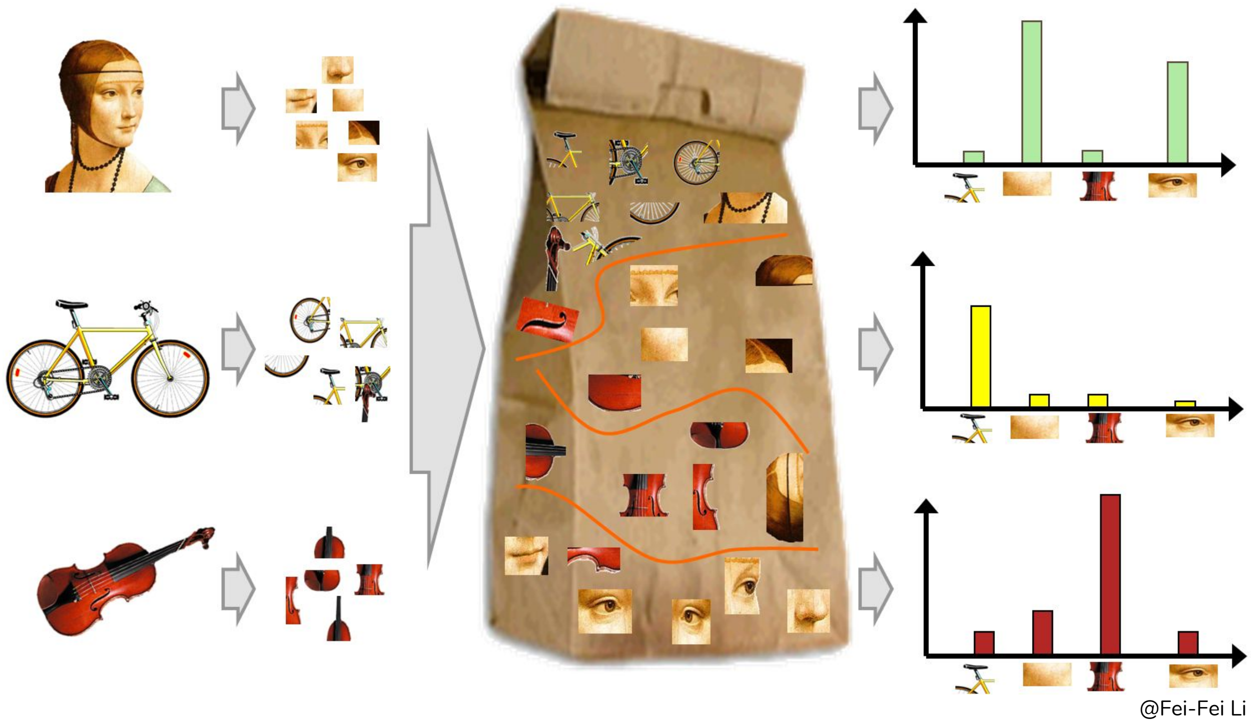
Local Feature Extraction

$$\begin{bmatrix} l_{1,1} & \dots & l_{1,N} \\ l_{2,1} & \dots & l_{2,N} \\ \vdots & & \vdots \\ l_{M,1} & \dots & l_{M,N} \end{bmatrix}$$

Mid-level Feature Extraction

Bag-of-Visual-Words (BoVW)

[Sivic and Zisserman, 2003; Csurka et al., 2004]



References
J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In: ICCV, 2003.
G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In: ECCV, 2004.

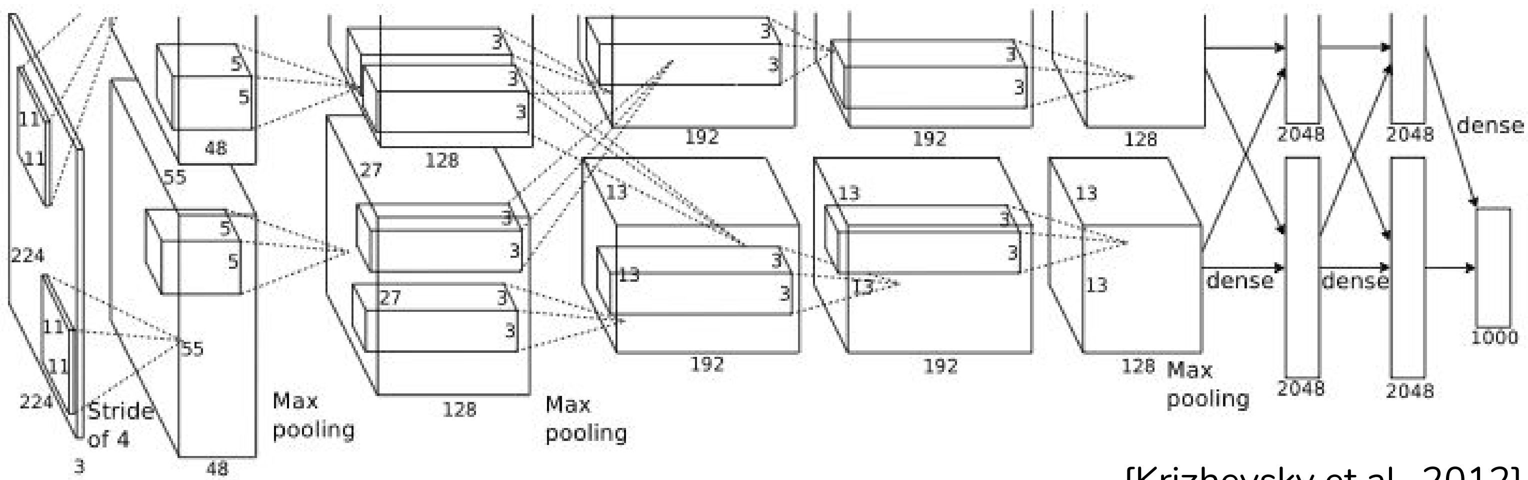
Mid-level Feature Extraction

BoVW-based Approaches

- Soft-assignment [van Gemert et al., 2010]
- Fisher Vector [Perronnin et al., 2010]
- VLAD [Jegou et al., 2010]
- Super-Vector Coding [Zhou et al., 2010]
- Spatial Fisher Vector [Krapac et al., 2011]
- Semi-Soft Coding [Liu et al., 2011]
- SSC [Oliveira et al., 2012]
- Compact VLAT [Negrel et al., 2012]
- BossaNova [Avila et al., 2013]
- BossaNova Video Descriptor [Caetano et al., 2016]

Deep Learning

Architecture of the ImageNet Challenge 2012 Winner

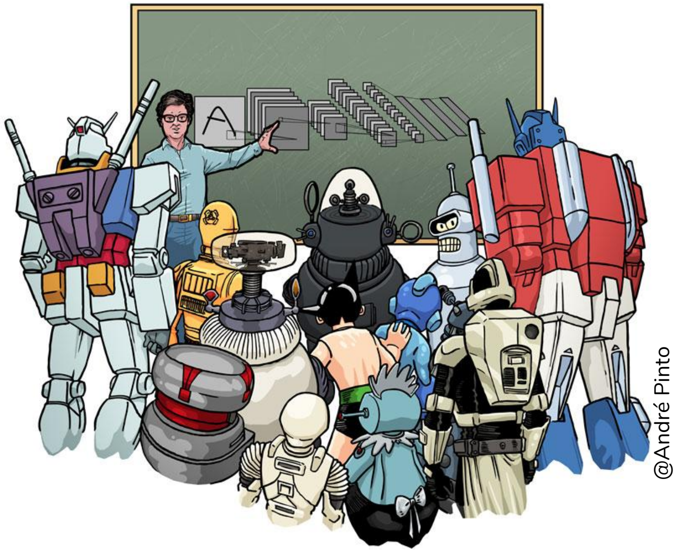


[Krizhevsky et al., 2012]

References
A Krizhevsky, I Sutskever, G. Hinton. ImageNet classification with deep convolutional neural networks. In: NIPS, 2012.

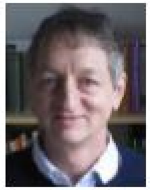
Deep Learning

- **Key Dates**
- **Convolutional Neural Networks (CNNs)**



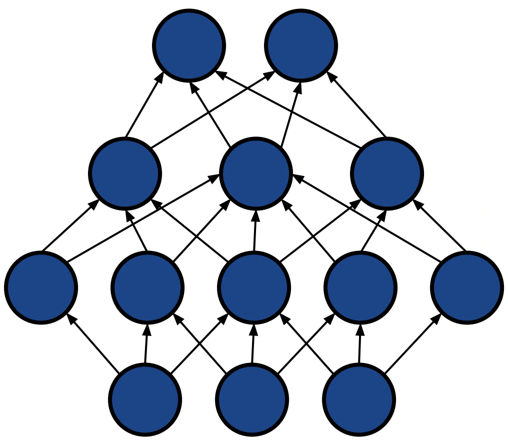
Neural network
Back propagation

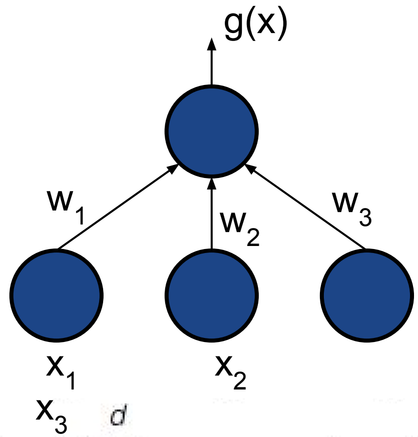
Nature

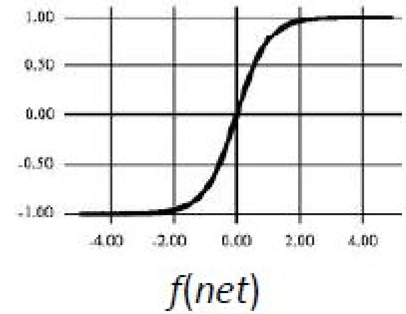


"Learning representations by back-propagating errors"

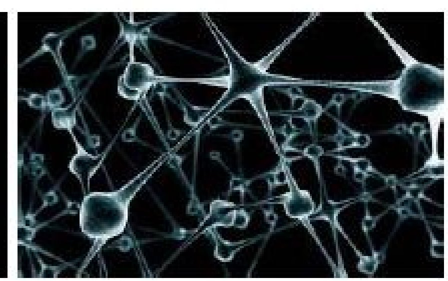
1986





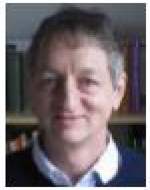


$$g(\mathbf{x}) = f\left(\sum_{i=1}^{d} x_i w_i + w_0\right) = f(\mathbf{w}^t \mathbf{x})$$

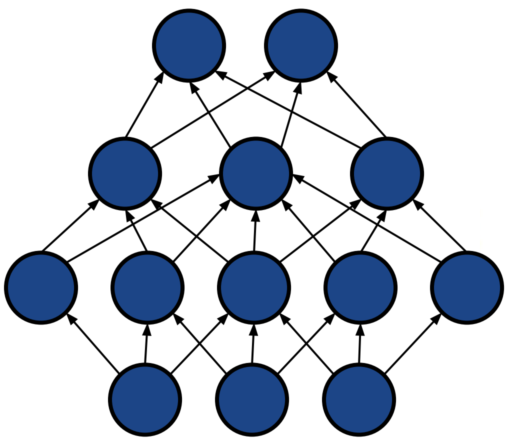


Neural network
Back propagation

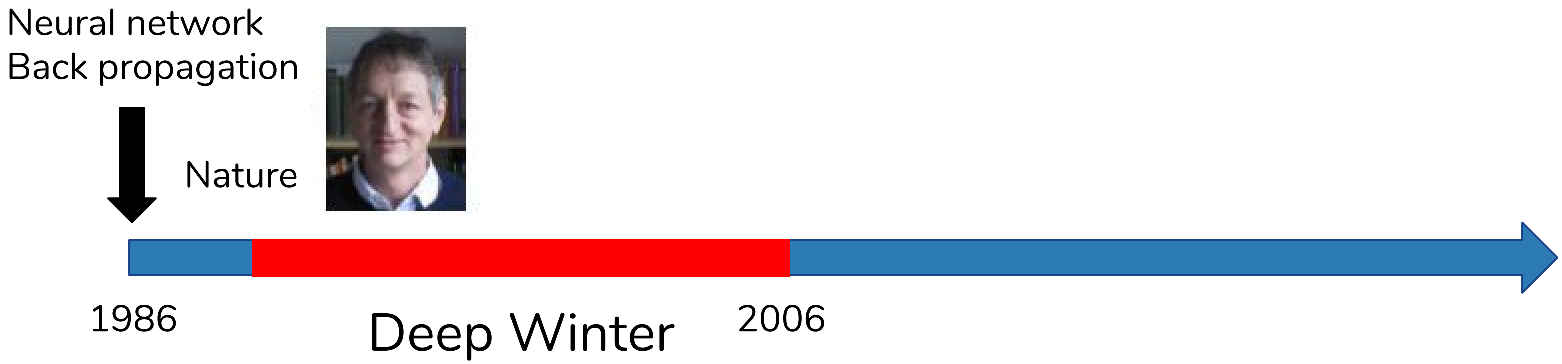
Nature



1986

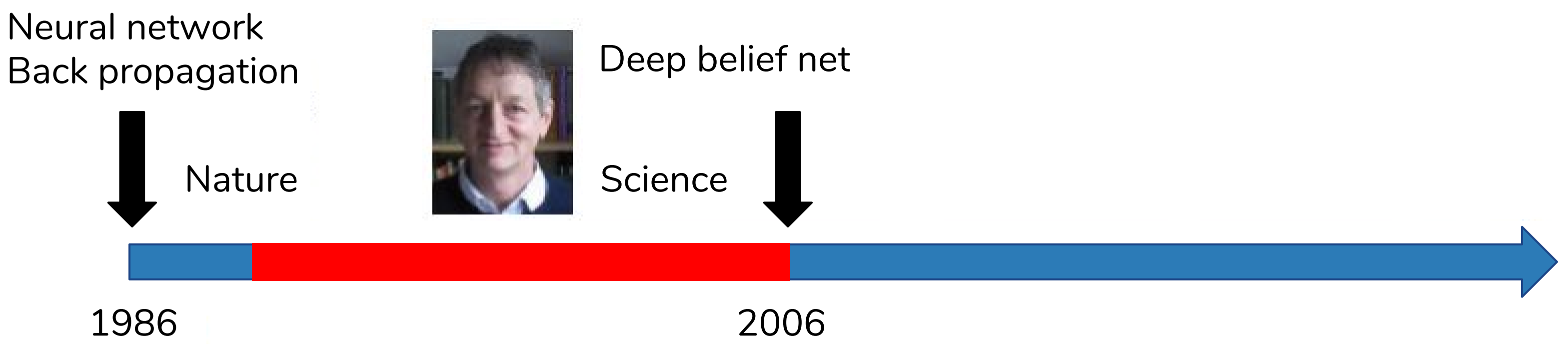


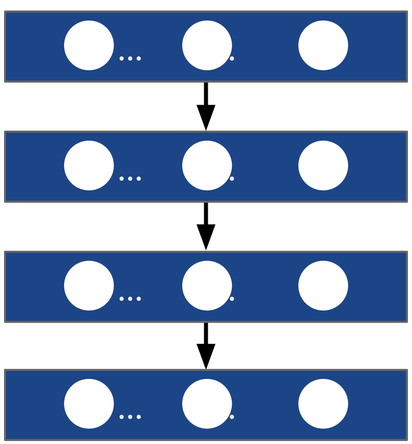
- Solve general learning problems
  - Tied with biological system
- But it is given up ...
  - Hard to train
  - Insufficient computational resources
  - Small training sets
  - Does not work well



- SVM
- Boosting
- Decision Tree
- KNN
- …

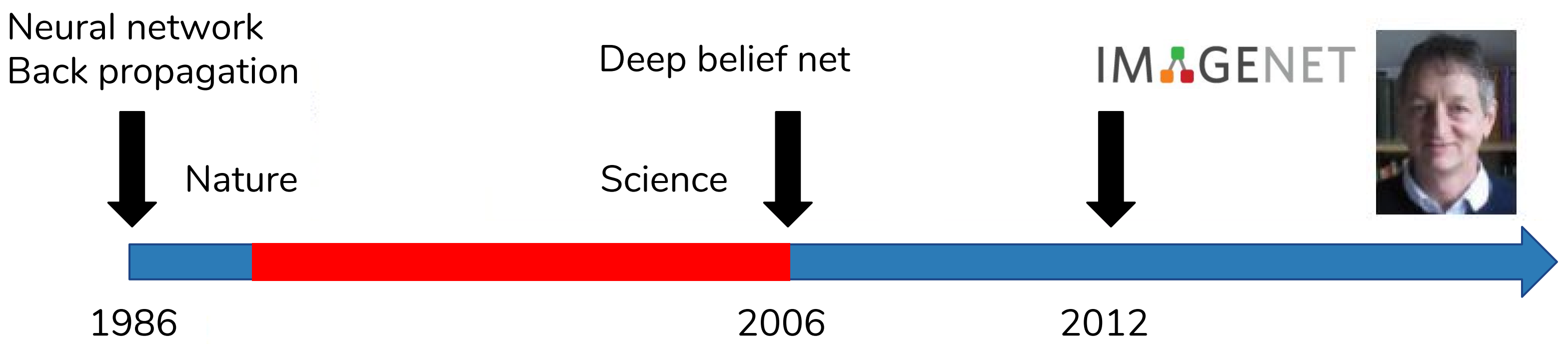
- Flat structures
- Loose tie with biological systems
- Specific methods for specific tasks
  - Hand crafted features (GMM-HMM, SIFT, LBP, HOG)





- Unsupervised & Layer-wised pre-training
- Better designs for modeling and training (normalization, nonlinearity, dropout)
- New development of computer architectures
    - GPU
    - Multi-core computer systems
- Large scale databases

Big Data!



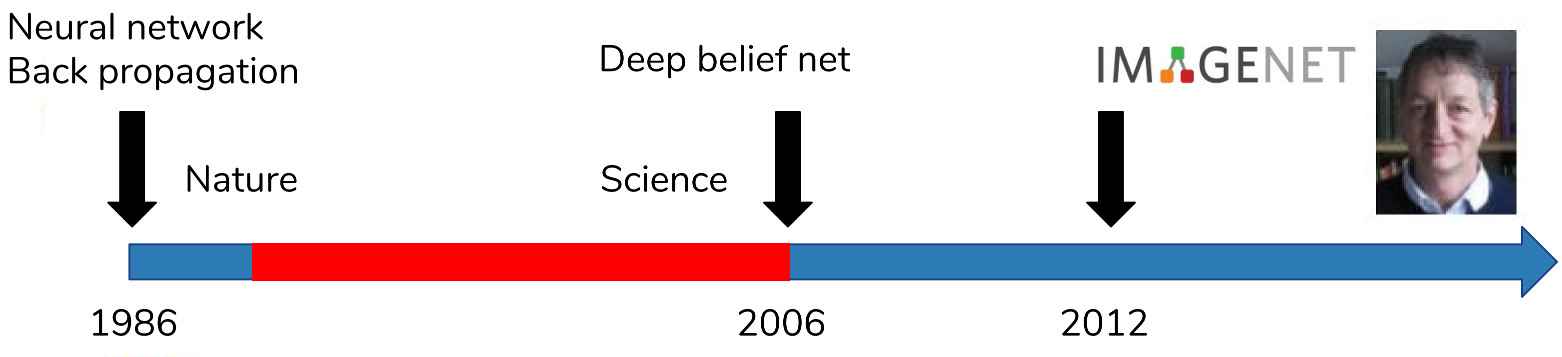
ILSVRC 2012 — Image Classification task

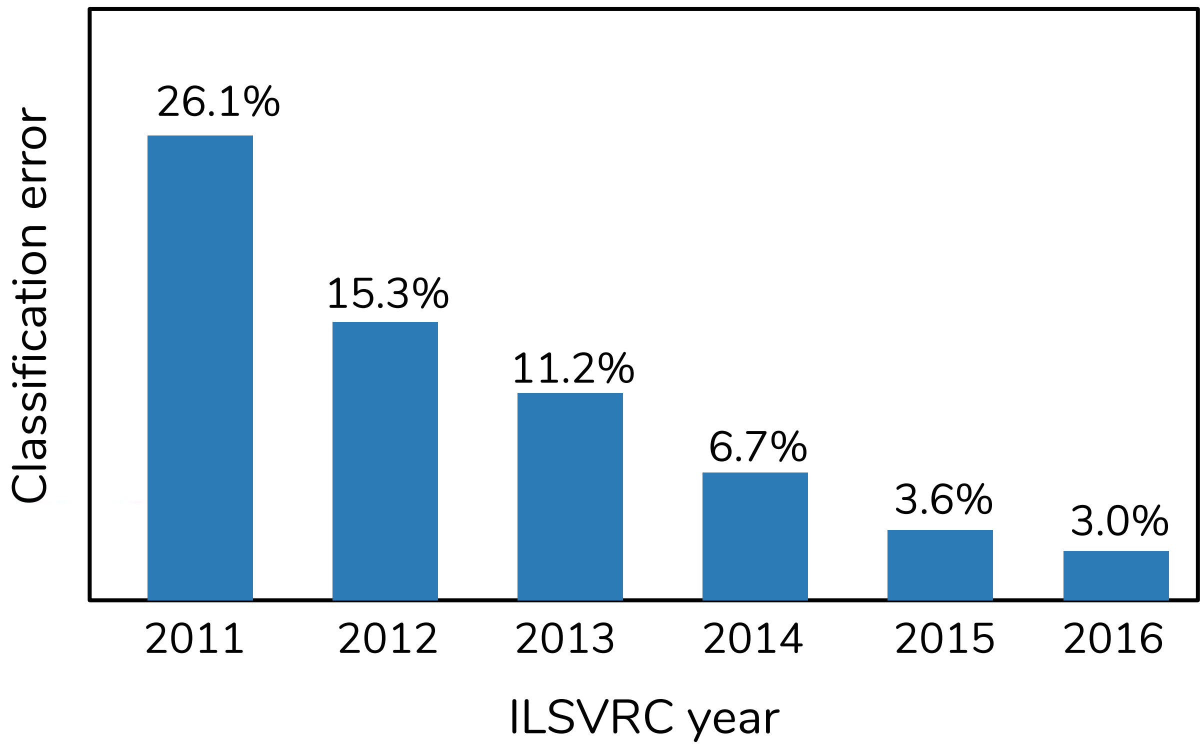
Rank	Name	Error Rate	Description
1	**U. Toronto**	15.3%	Deep Learning
2	U. Tokyo	26.2%	Hand-crafted features and learning models
3	U. Oxford	26.9%	
4	Xerox/INRIA	27.0%	

Object recognition over 1,000,000 images and 1,000 categories (2 GPU)

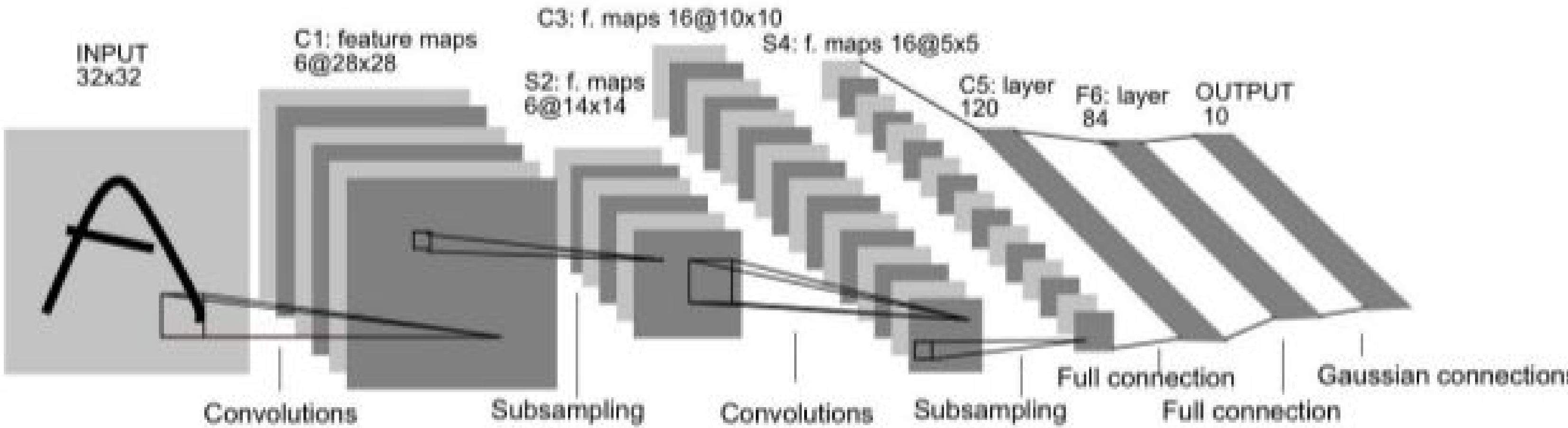
References

A Krizhevsky, I Sutskever, G. Hinton. ImageNet classification with deep convolutional neural networks. In: NIPS, 2012.


Neural network
Back propagation
Nature
Deep belief net
Science
IMAGENET
1986
2006
2012


26.1%
15.3%
11.2%
6.7%
3.6%
3.0%
Classification error
2011
2012
2013
2014
2015
2016
ILSVRC year

Convolutional Neural Networks (CNNs)



(Filtering) (Coding) (Pooling) (Filtering) (Pooling) (Classification)

[LeCun et al., 1998]

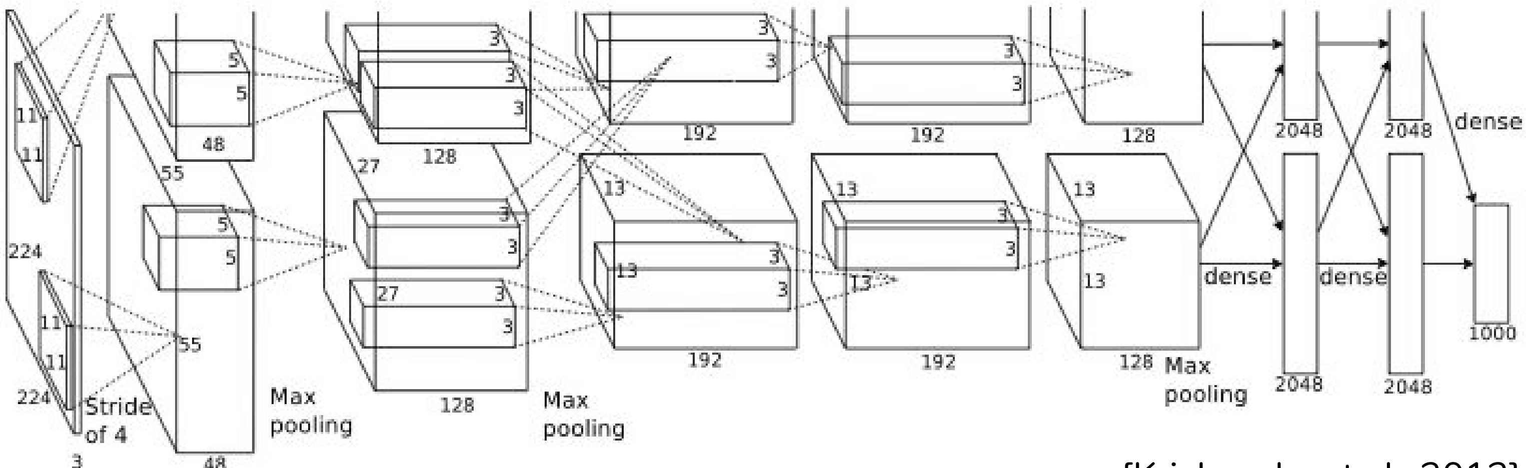
- **Convolution** uses local weights shared across the whole image
- **Pooling** shrinks the spatial dimensions

References

Y. LeCun, L. Bottou, Y. Bengio. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998

Large CNNs

Architecture of the ImageNet Challenge 2012 Winner

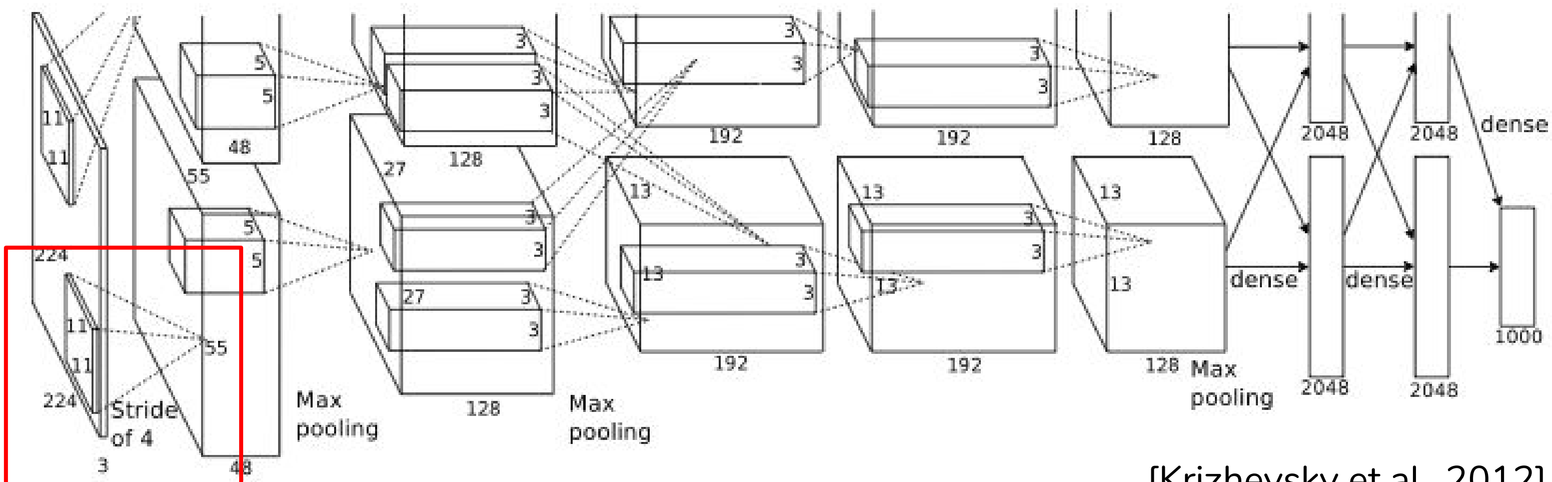


[Krizhevsky et al., 2012]

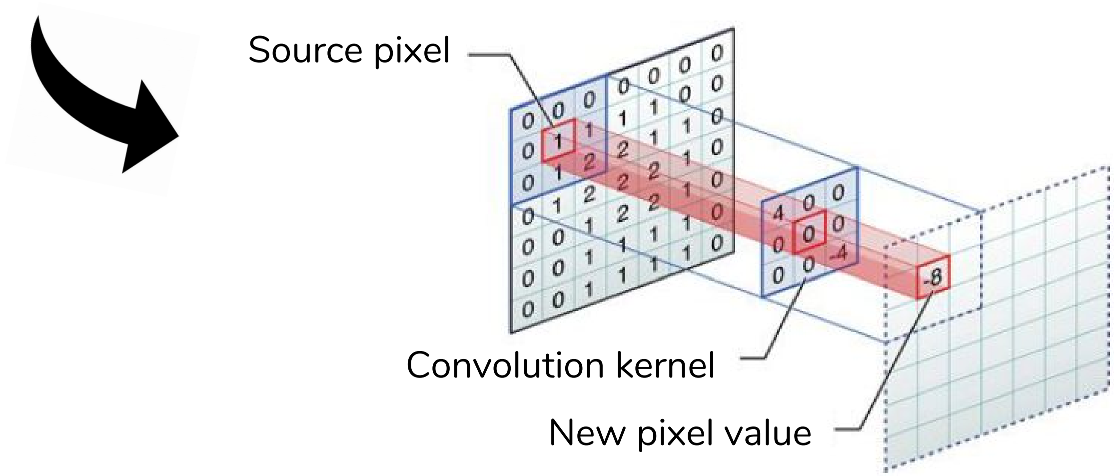
- Same model as LeCun but:
  - Bigger model (8 layers)
  - More data (10^6 vs. 10^3 images)
  - GPU implementation (50x speedup over CPU)
  - Better regularization (DropOut)

References
A Krizhevsky, I Sutskever, G. Hinton. ImageNet classification with deep convolutional neural networks. In: NIPS, 2012.

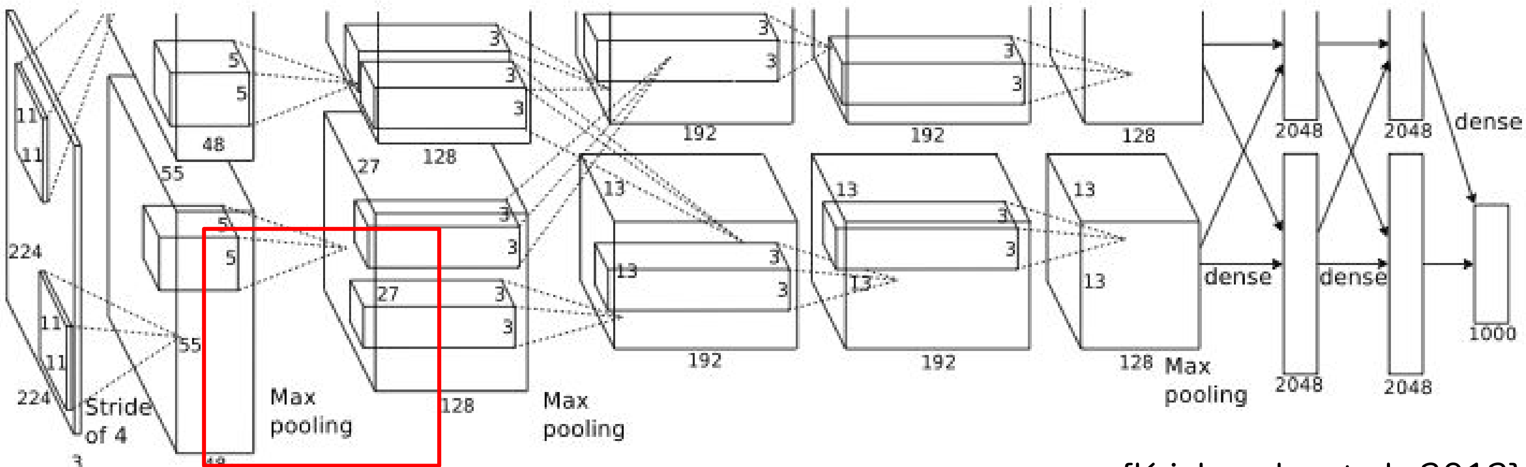
Large CNNs



[Krizhevsky et al., 2012]



Large CNNs



[Krizhevsky et al., 2012]

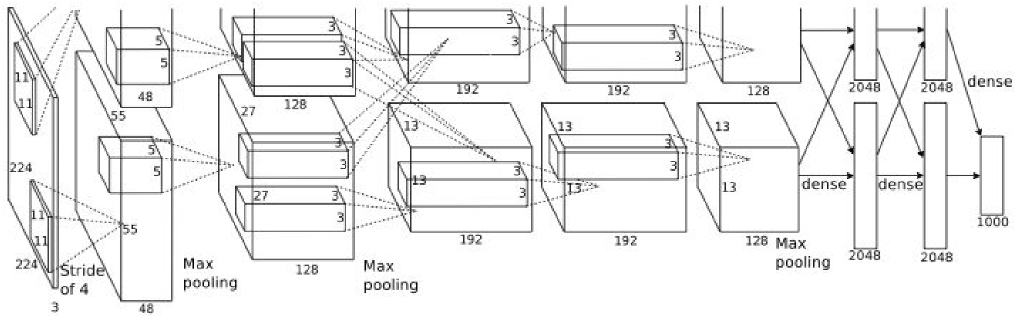
1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

Max pooling with 2x2 filters and stride 2

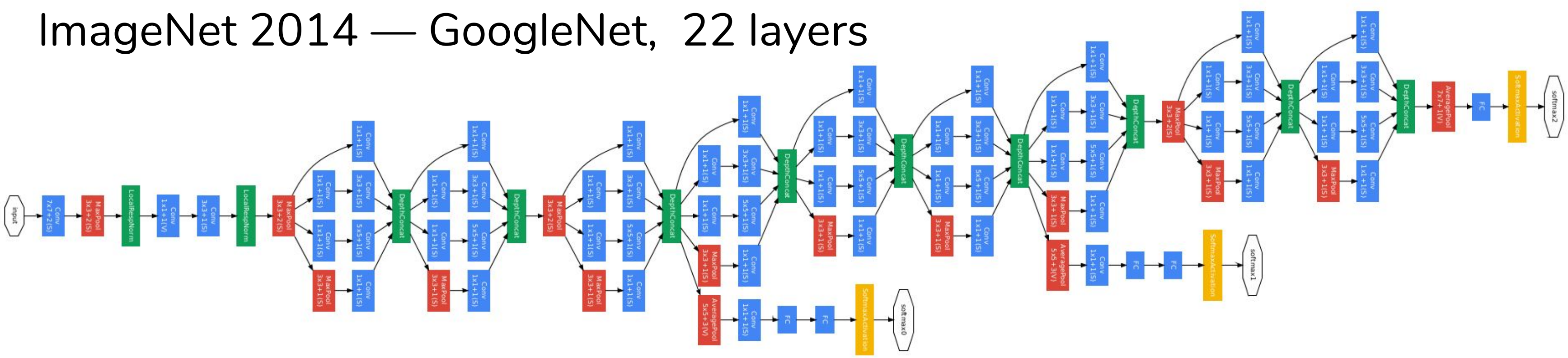
6	8
3	4

Very Large CNNs

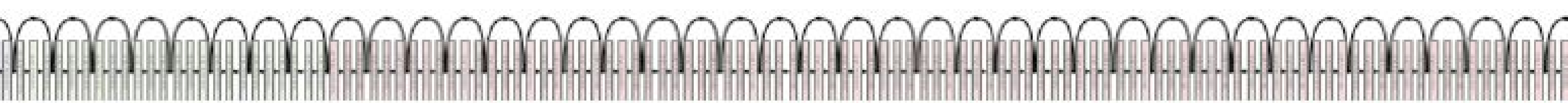
ImageNet 2012 — AlexNet, 8 layers



ImageNet 2014 — GoogleNet, 22 layers



ImageNet 2015 — ResNet, 152 layers



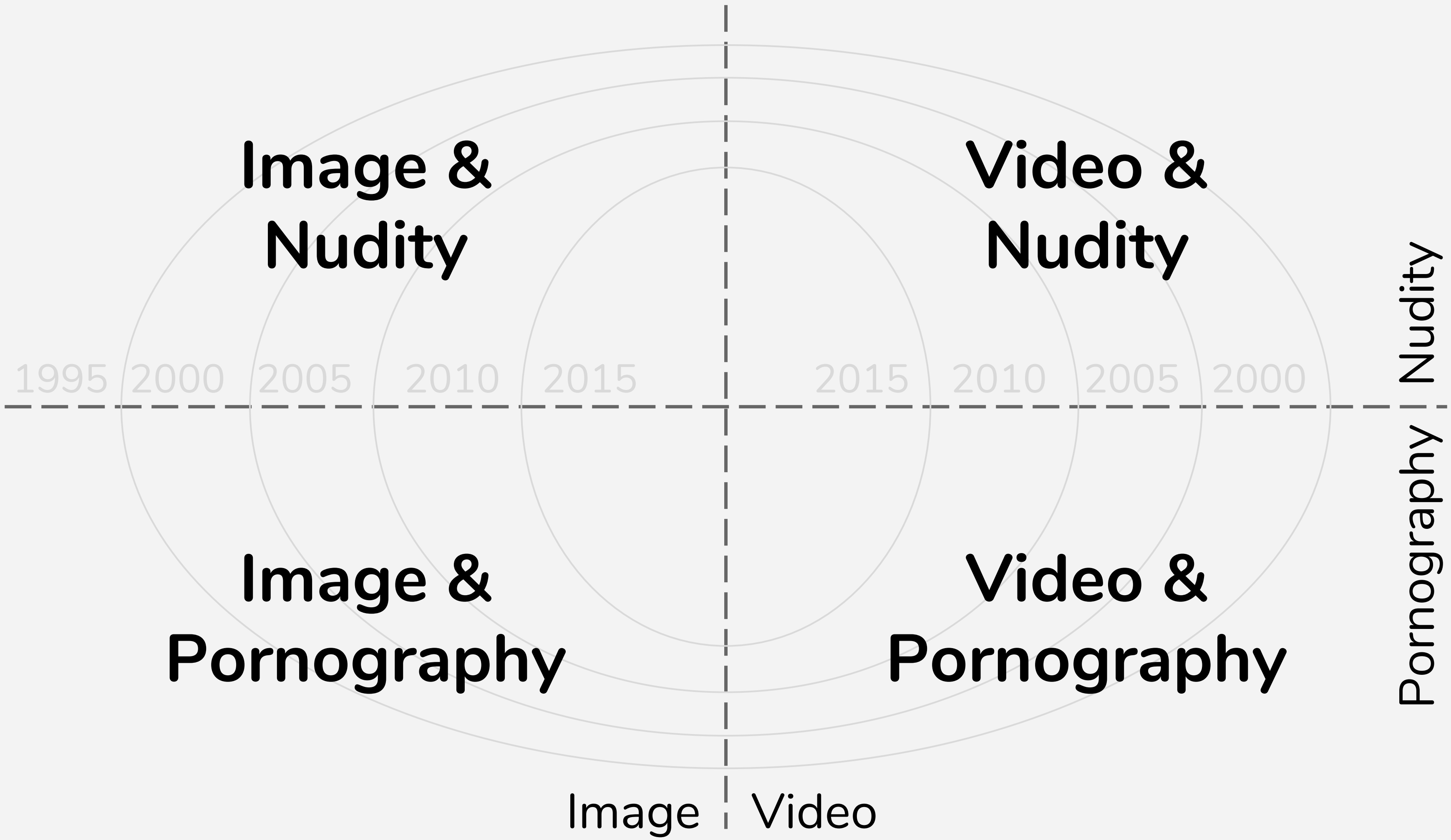
Our Applications

Sensitive Media Analysis
(Pornography)

Literature

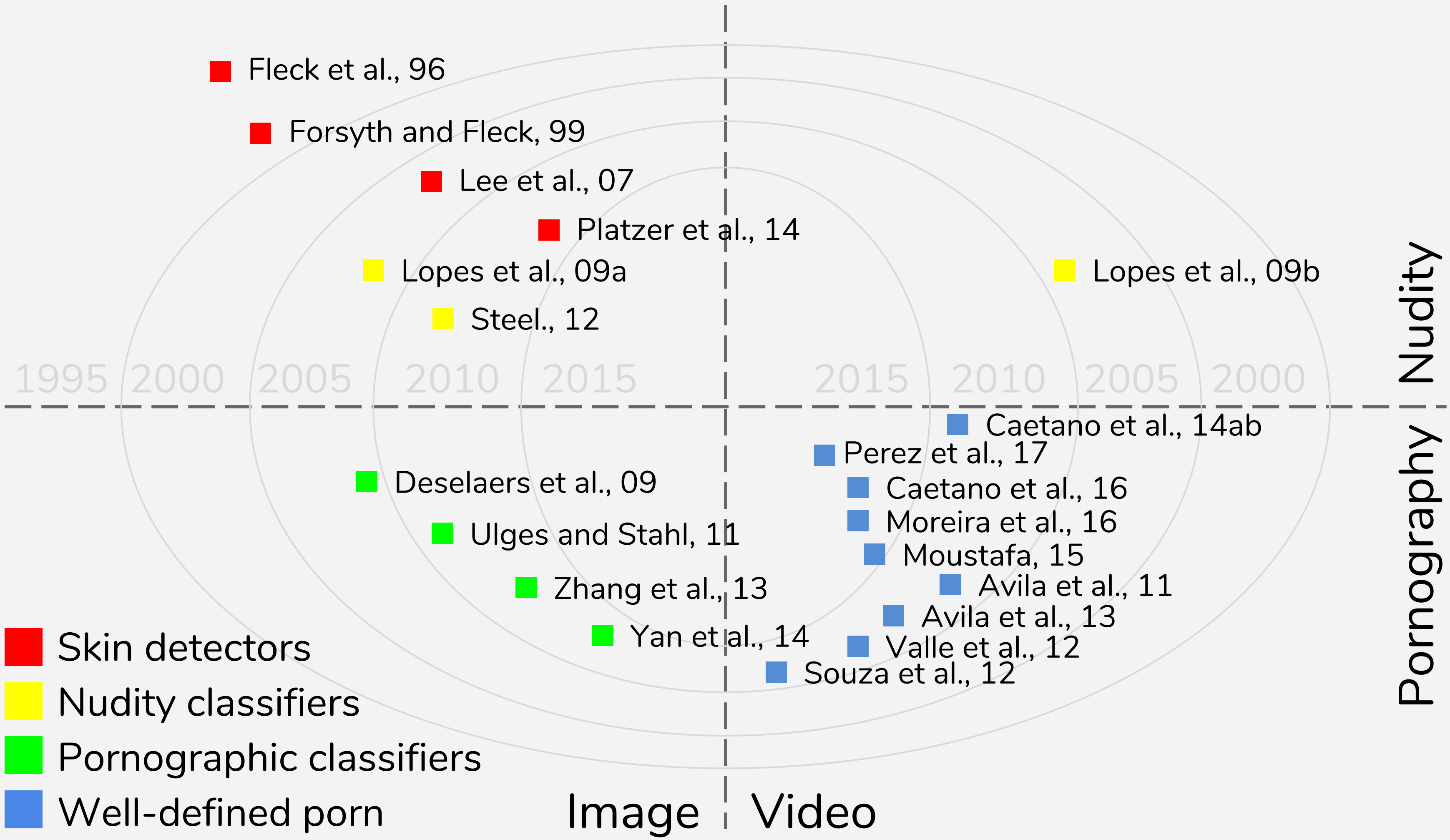
(Pornography)

Pornography Classification



@Daniel Moreira 2016

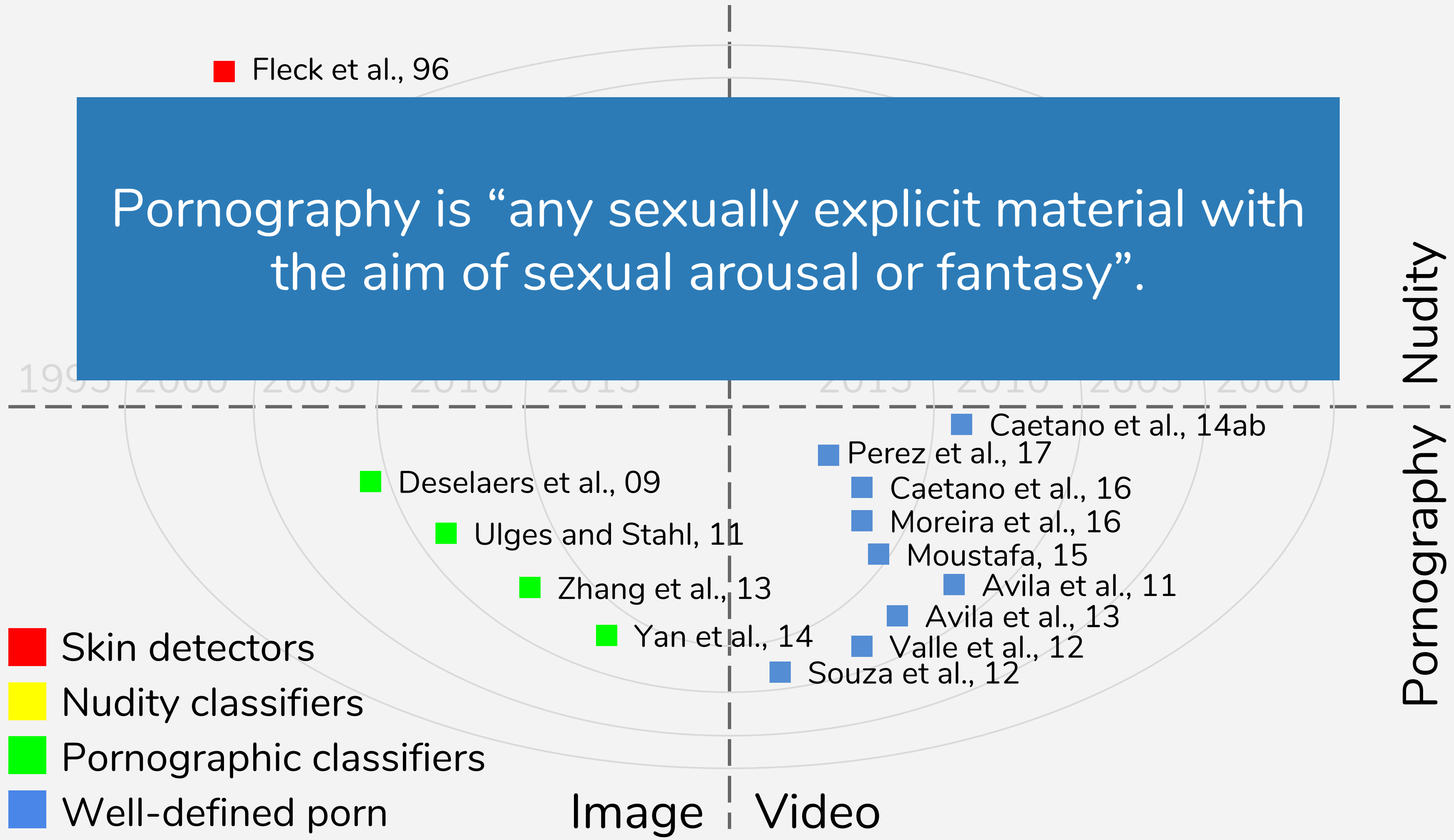
Pornography Classification



@Daniel Moreira 2016

Pornography Classification

Pornography is “any sexually explicit material with the aim of sexual arousal or fantasy”.



@Daniel Moreira 2016

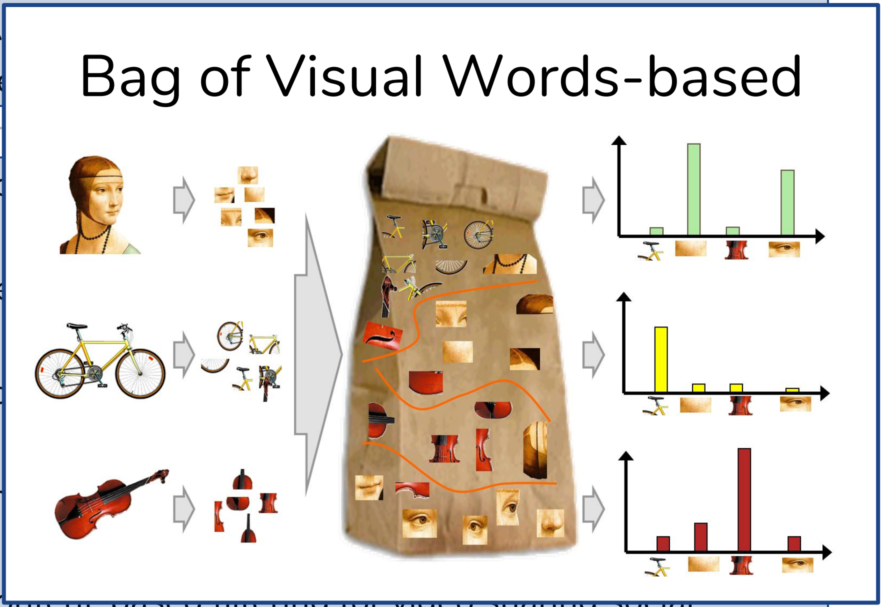
Short et al. A review of internet pornography use research: methodology and content from the past 10 years. Cyberpsychology, Behavior, and Social Networking, 2012

Pornography Classification

- M. Perez, **S. Avila**, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, Video pornography detection through deep learning techniques and motion information. Neurocomputing, 2017
- D. Moreira, **S. Avila**, M. Perez, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, Pornography classification: The hidden clues in video space-time, FSI, 2016
- C. Caetano, **S. Avila**, W. Schwartz, S. Guimarães, A. Araújo. A mid-Level video representation based on binary descriptors: A case study for pornography detection. Neurocomputing, 2016
- M. Moustafa. Applying deep learning to classify pornographic images and videos. In: 7th PSIVT, 2015
- C. Caetano, **S. Avila**, S. Guimarães, A. Araújo. Pornography detection using BossaNova video descriptor. In: 22nd EUSIPCO, 2014
- C. Caetano, **S. Avila**, S. Guimarães, A. Araújo. Representing local binary descriptors with BossaNova for visual recognition. In: 29th ACM SAC, 2014
- **S. Avila**, N. Thome, M. Cord, E. Valle, A. Araújo. Pooling in image representation: the visual codeword point of view. CVIU, 2013
- F. Souza, E. Valle, G. Cámara-Chávez, A. Araújo. An evaluation on color invariant based local spatiotemporal features for action recognition. In: 25th SIBGRAPI, 2012
- E. Valle, **S. Avila**, F. Souza, M. Coelho, A.. Araújo. Content-based filtering for video sharing social networks. In: 12th SBSeg, 2012
- **S. Avila**, N. Thome, M. Cord, E. Valle, A. Araújo. BOSSA: Extended BoW formalism for image classification. In: 18th ICIP, 2011

Pornography Classification

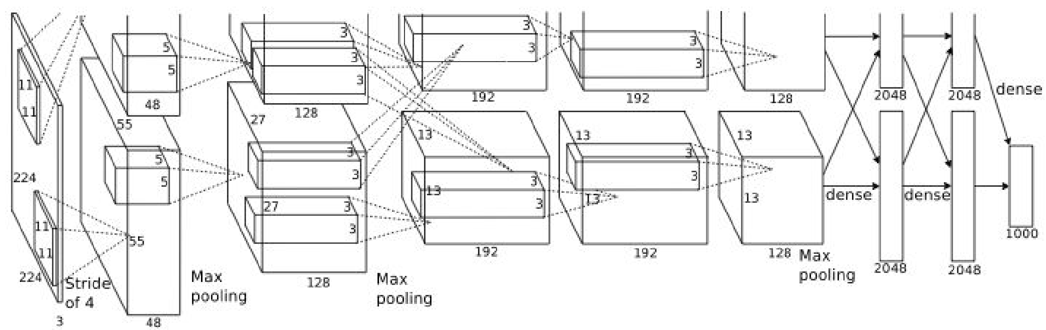
- M. Perez, **S. Avila**, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, Video pornography detection through deep learning techniques and motion information. Neurocomputing, 2017
- D. Moreira, **S. Avila**, M. Perez, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, Pornography classification: The hidden clues in video space-time, FSI, 2016
- C. Caetano, **S. Avila**, W. Schwartz, S. Guimarães, A
binary descriptors: A case study for pornography de
- M. Moustafa. Applying deep learning to classify por
- C. Caetano, **S. Avila**, S. Guimarães, A. Araújo. Porno
descriptor. In: 22nd EUSIPCO, 2014
- C. Caetano, **S. Avila**, S. Guimarães, A. Araújo. Repre
visual recognition. In: 29th ACM SAC, 2014
- **S. Avila**, N. Thome, M. Cord, E. Valle, A. Araújo. Poo
point of view. CVIU, 2013
- F. Souza, E. Valle, G. Cámara-Chávez, A. Araújo. Ar
spatiotemporal features for action recognition. In: 2
- E. Valle, **S. Avila**, F. Souza, M. Coelho, A.. Araújo. Content-based filtering for video sharing social networks. In: 12th SBSeg, 2012
- **S. Avila**, N. Thome, M. Cord, E. Valle, A. Araújo. BOSSA: Extended BoW formalism for image classification. In: 18th ICIP, 2011



Pornography Classification

- M. Perez, **S. Avila**, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, Video pornography detection through deep learning techniques and motion information. Neurocomputing, 2017
- D. Moreira, **S. Avila**, M. Perez, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha, Pornography classification: The hidden clues in video space-time, FSI, 2016
- C. Caetano, **S. Avila**, W. Schwartz, S. Guimarães, A. Araújo. A mid-Level video representation based on binary descriptors: A case study for pornography detection. Neurocomputing, 2016
- M. Moustafa. Applying deep learning to classify pornographic images and videos. In: 7th PSIVT, 2015
- C. Caetano, **S. Avila**, S. Guimarães, A. Araújo. Pornography detection using BossaNova video descriptor. In: 22nd EUSIPCO, 201
- C. Caetano, **S. Avila**, S. Guimarães visual recognition. In: 29th ACM S
- **S. Avila**, N. Thome, M. Cord, E. Val point of view. CVIU, 2013
- F. Souza, E. Valle, G. Cámara-Cháv spatiotemporal features for action
- E. Valle, **S. Avila**, F. Souza, M. Coel networks. In: 12th SBSeg, 2012
- **S. Avila**, N. Thome, M. Cord, E. Valle, A. Araújo. BOSSA: Extended BoW formalism for image classification. In: 18th ICIP, 2011

Deep Learning-based



Pornography Classification

Off-the-shelf Solutions

Media Detective

SnitchPlus

Belkasoft
forensics made easier

PORN DETECTION STICK

K9 Web Protection
from Blue Coat

Net Nanny

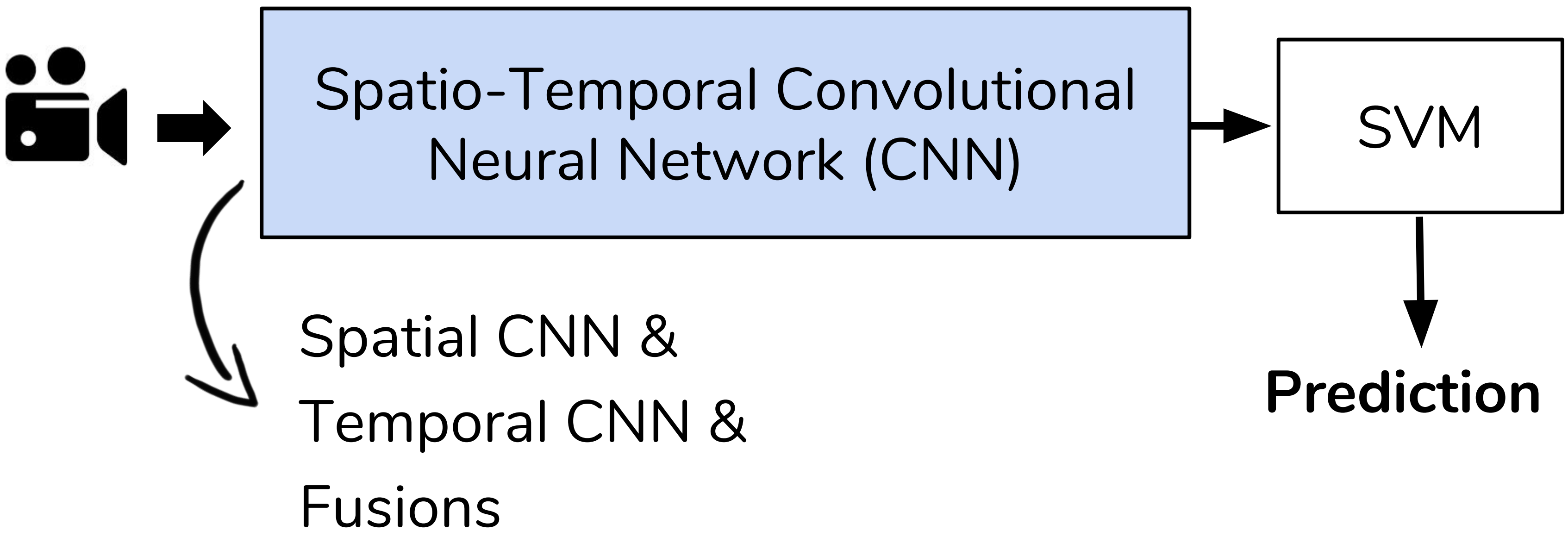
CyberPatrol
protecting people online

Our Solution

(Pornography)

Pornography Classification

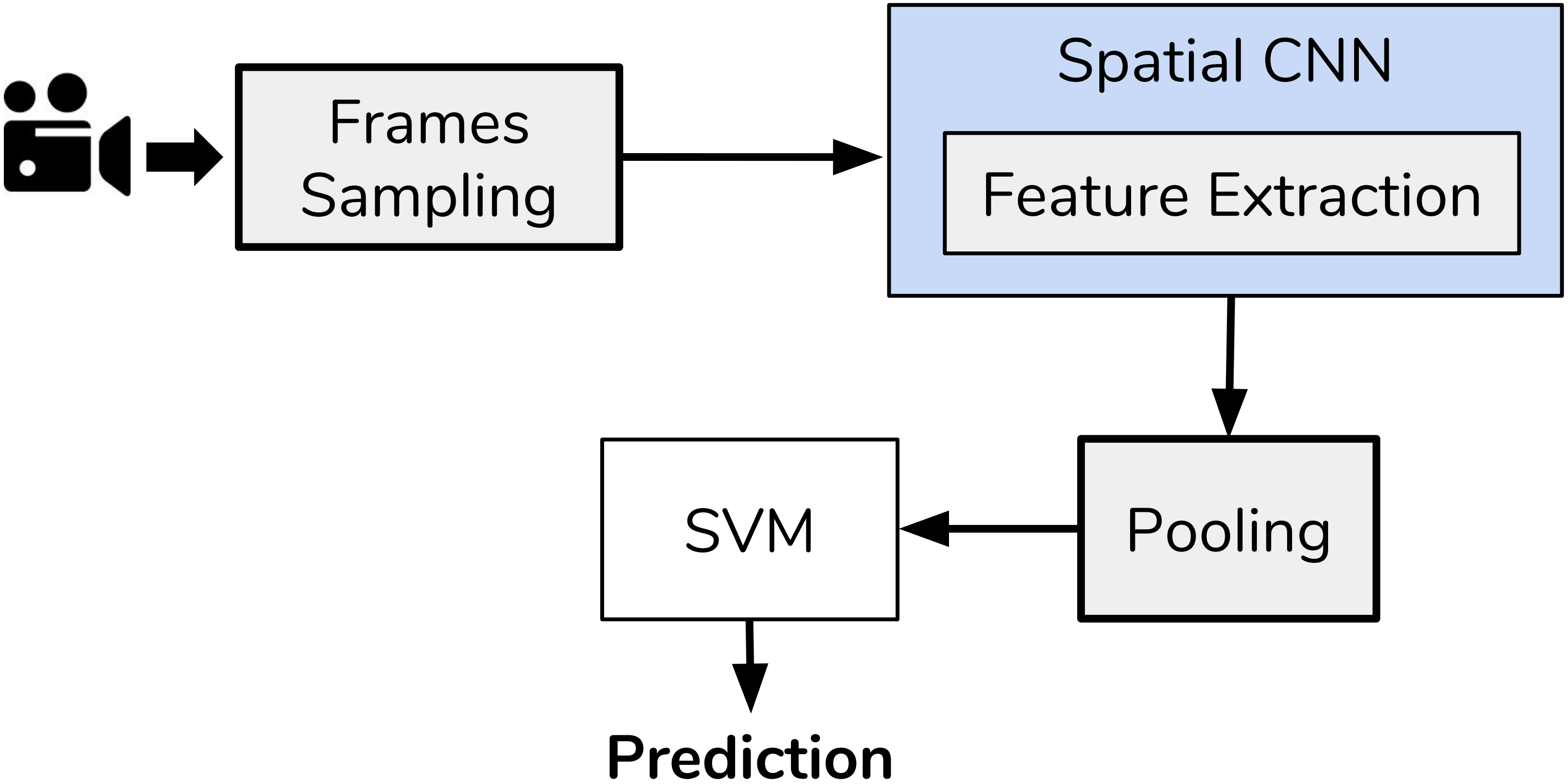
Deep Learning-based Approach [Perez et al., 2017]



M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, A. Rocha. Video pornography detection through deep learning techniques and motion information. Neurocomputing, 2017.

Pornography Classification

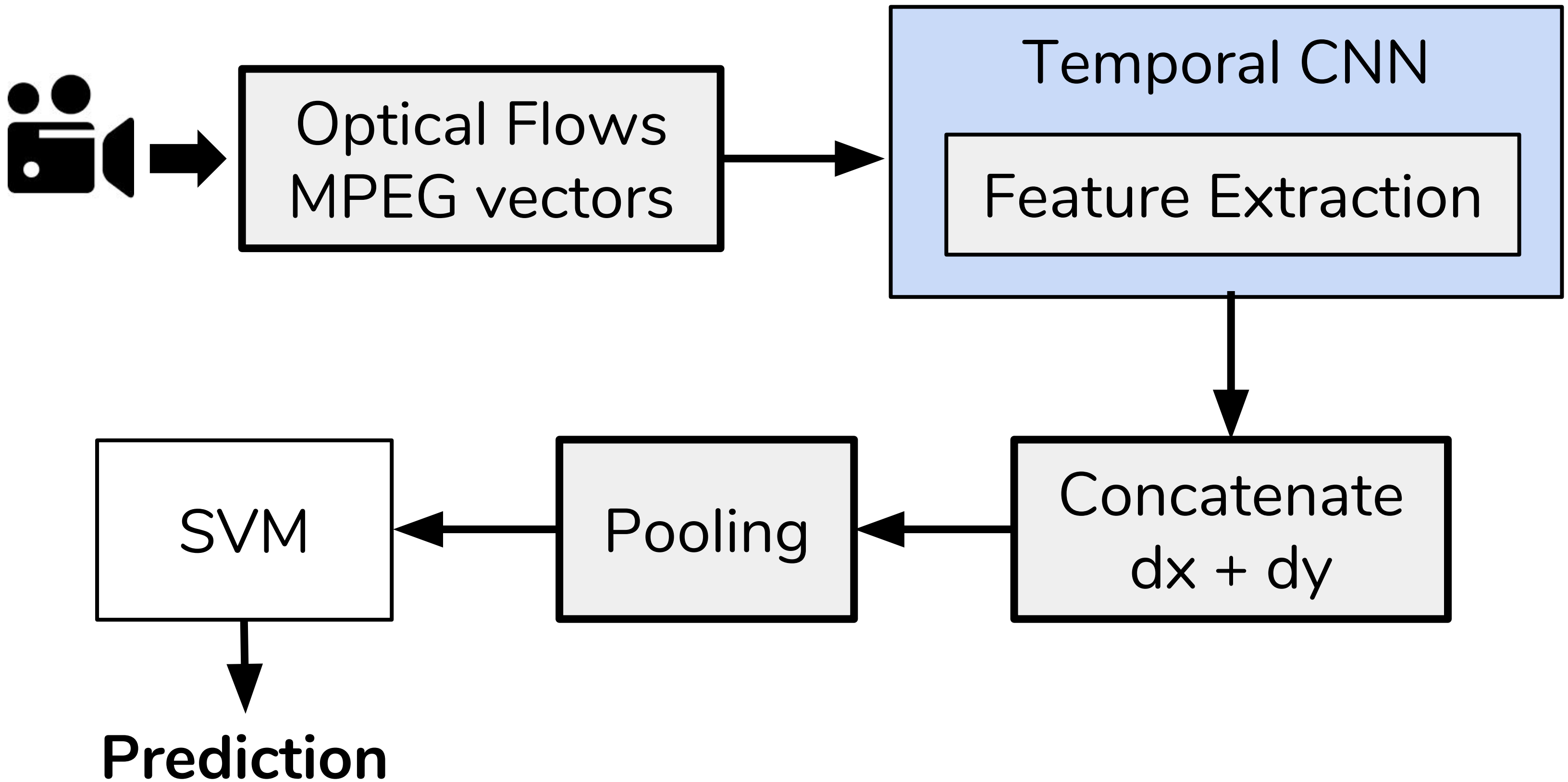
Deep Learning-based Approach [Perez et al., 2017]



M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, A. Rocha. Video pornography detection through deep learning techniques and motion information. Neurocomputing, 2017.

Pornography Classification

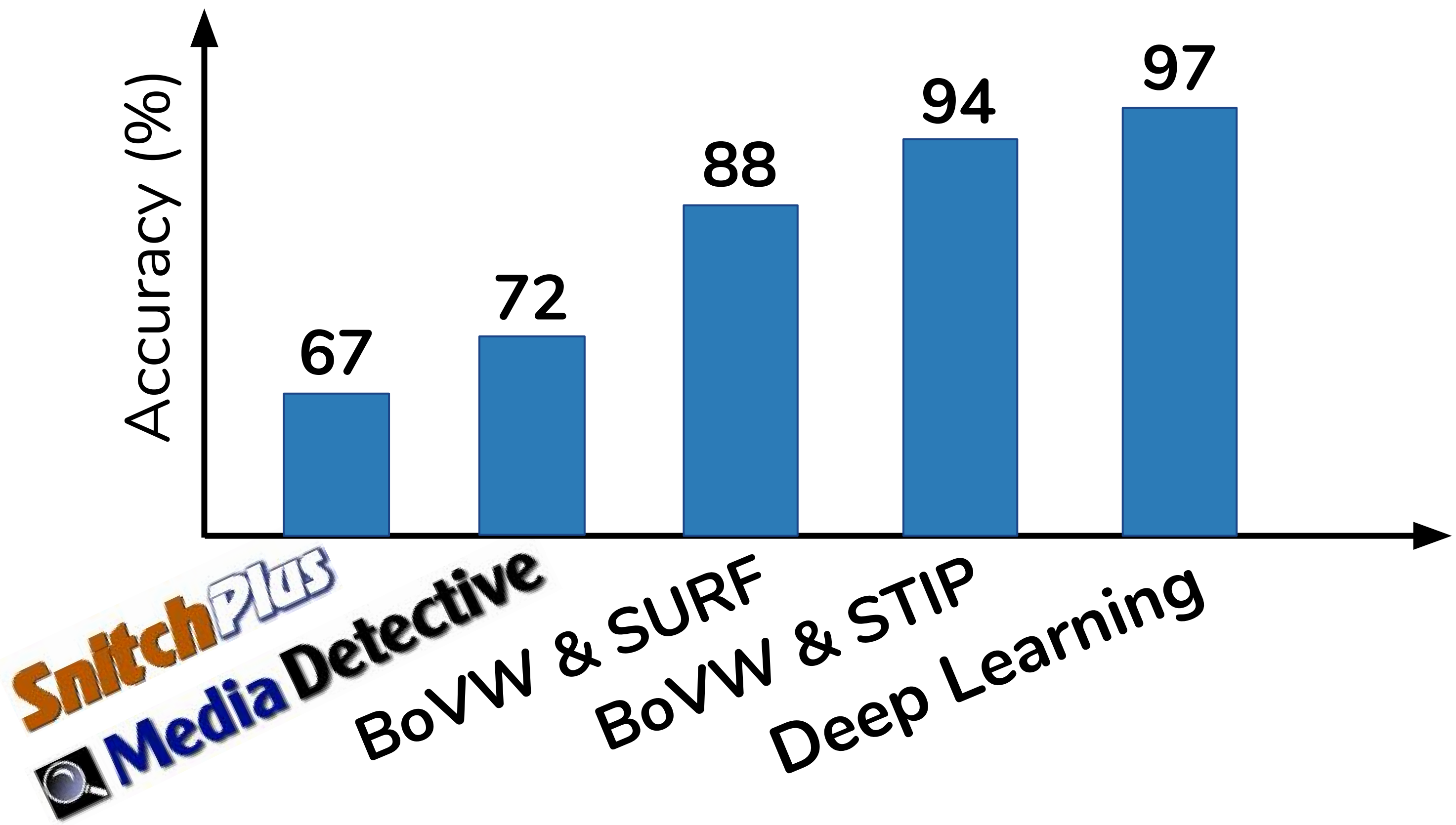
Deep Learning-based Approach [Perez et al., 2017]



M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, A. Rocha. Video pornography detection through deep learning techniques and motion information. Neurocomputing, 2017.

Pornography Classification

2,000 videos
140 hours



Our Applications

Sensitive Media Analysis
(Child Pornography)

Sexual predators can hide in your child’s smartphone.

Literature
(Child Pornography)

What literature is doing

Hash Matching

Oliveira & Silva 2009
Vrubel 2011
Hurley et al. 2013

Skin Detection

Polastro & Eleuterio 2010
Islam et al. 2010
Sae-Bae et al. 2014
Chatzis et al. 2016

Bag of Visual Words

Ulges et al. 2011
Carvalho et al. 2012
Schulze et al. 2014
Vitorino et al. 2016

What Forensic Tools are doing

Hash Matching

Microsoft
PhotoDNA
INACT

Skin Detection

NuDetective
(Federal Police)
Media Detective
Snitch Plus

Bag of Visual Words

✖

Our Solution

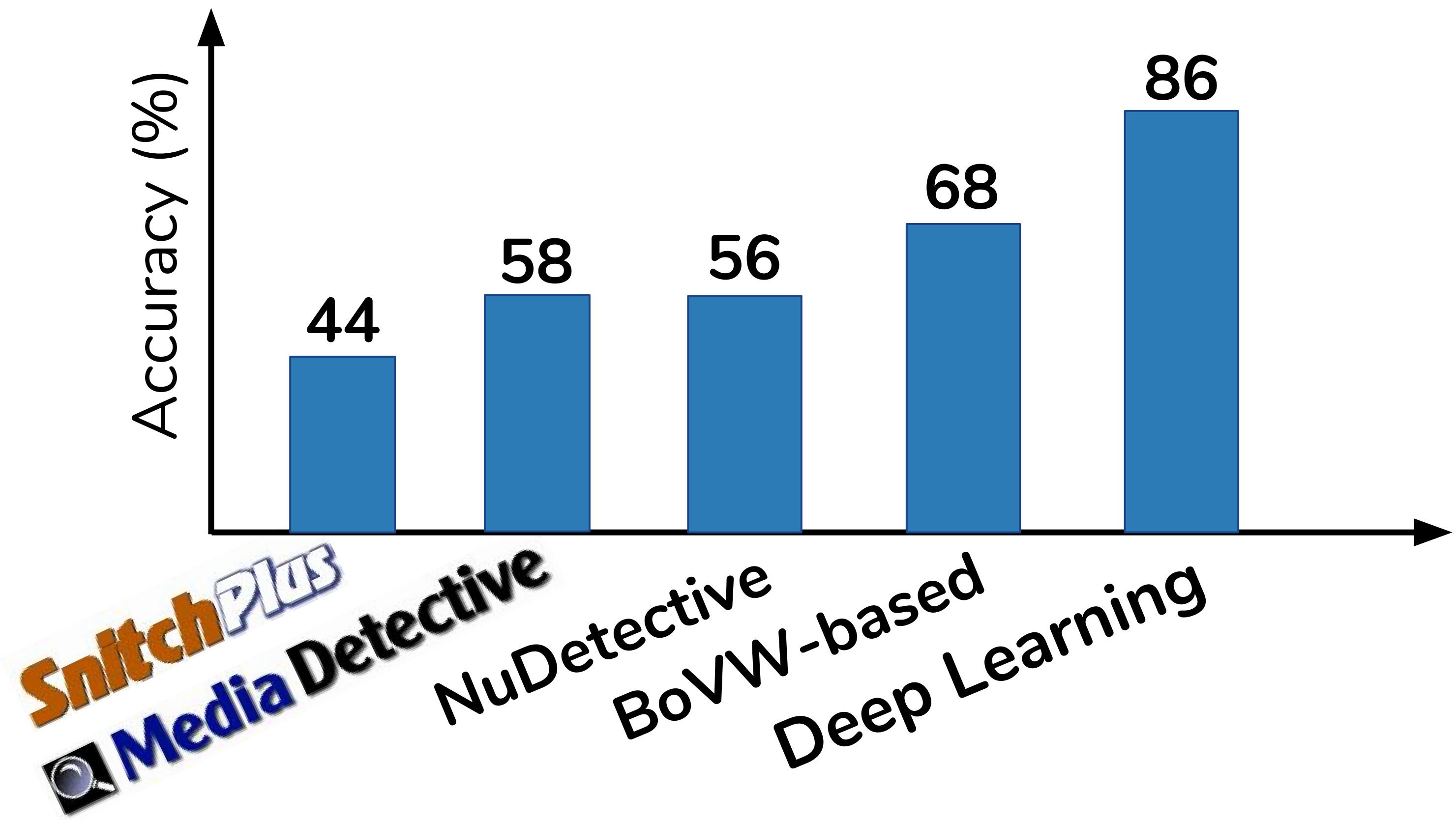
(Child Pornography)

Child Pornography Classification

- **Bag of Visual Words**-based solution
  - SURF + BossaNova + SVM
- **Deep Learning**-based solution
  - Transfer learning + Fine tuning + SVM

- **Dataset:** 60k images

Child Pornography Classification



Our Applications

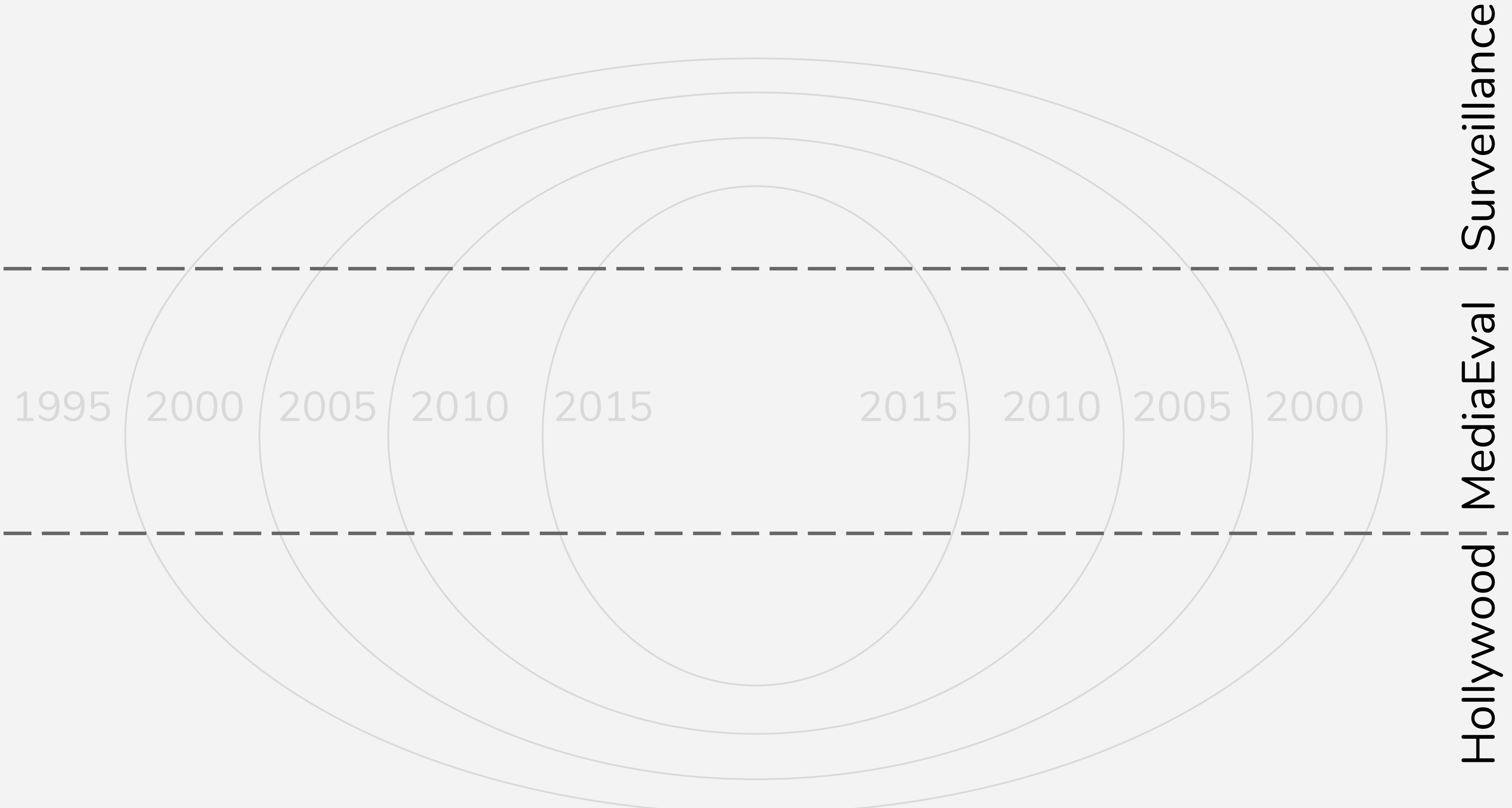
Sensitive Media Analysis
(Violence)


FOX
TOLEDO
10:01 69°

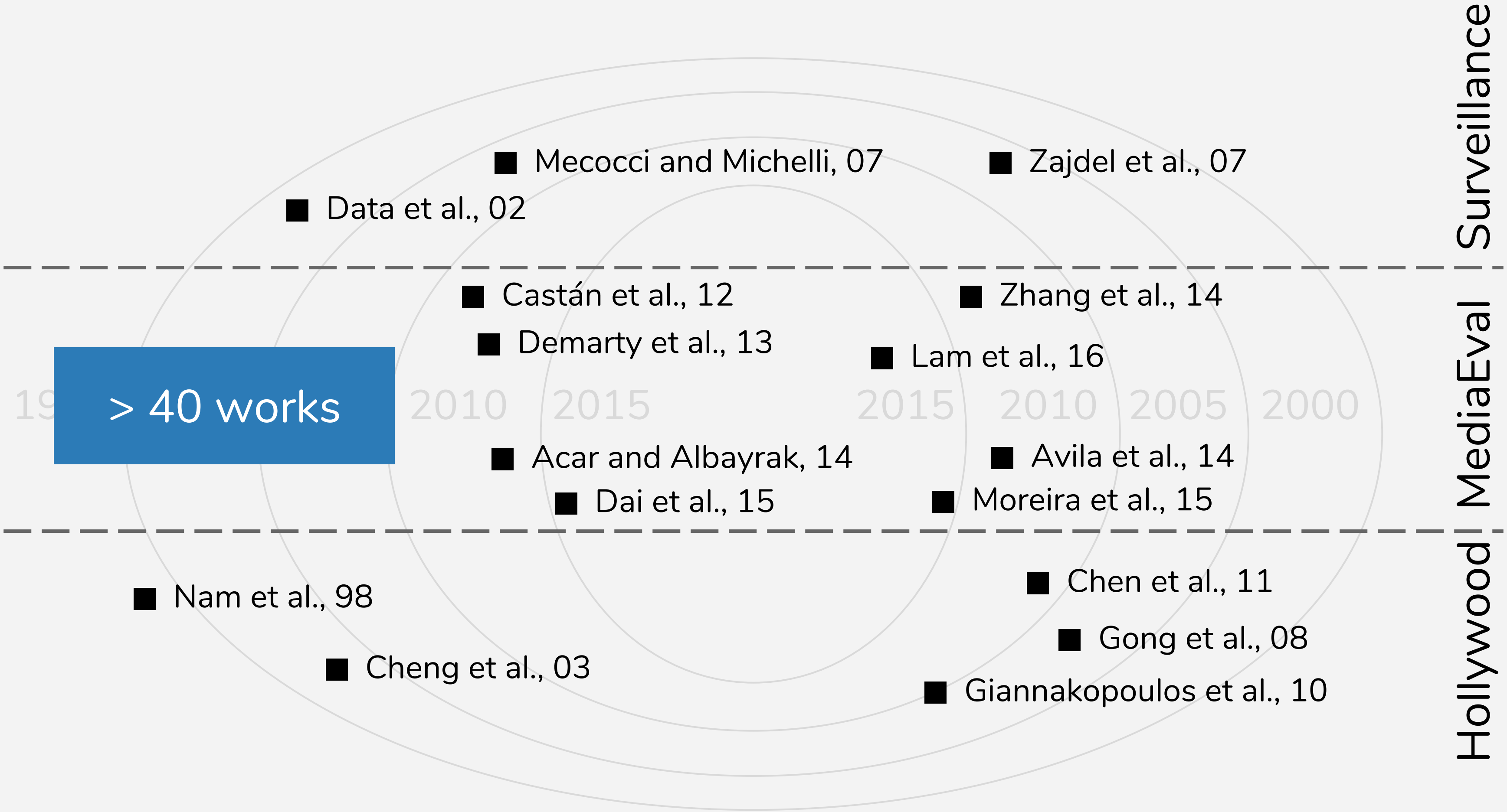
Literature

(Violence)

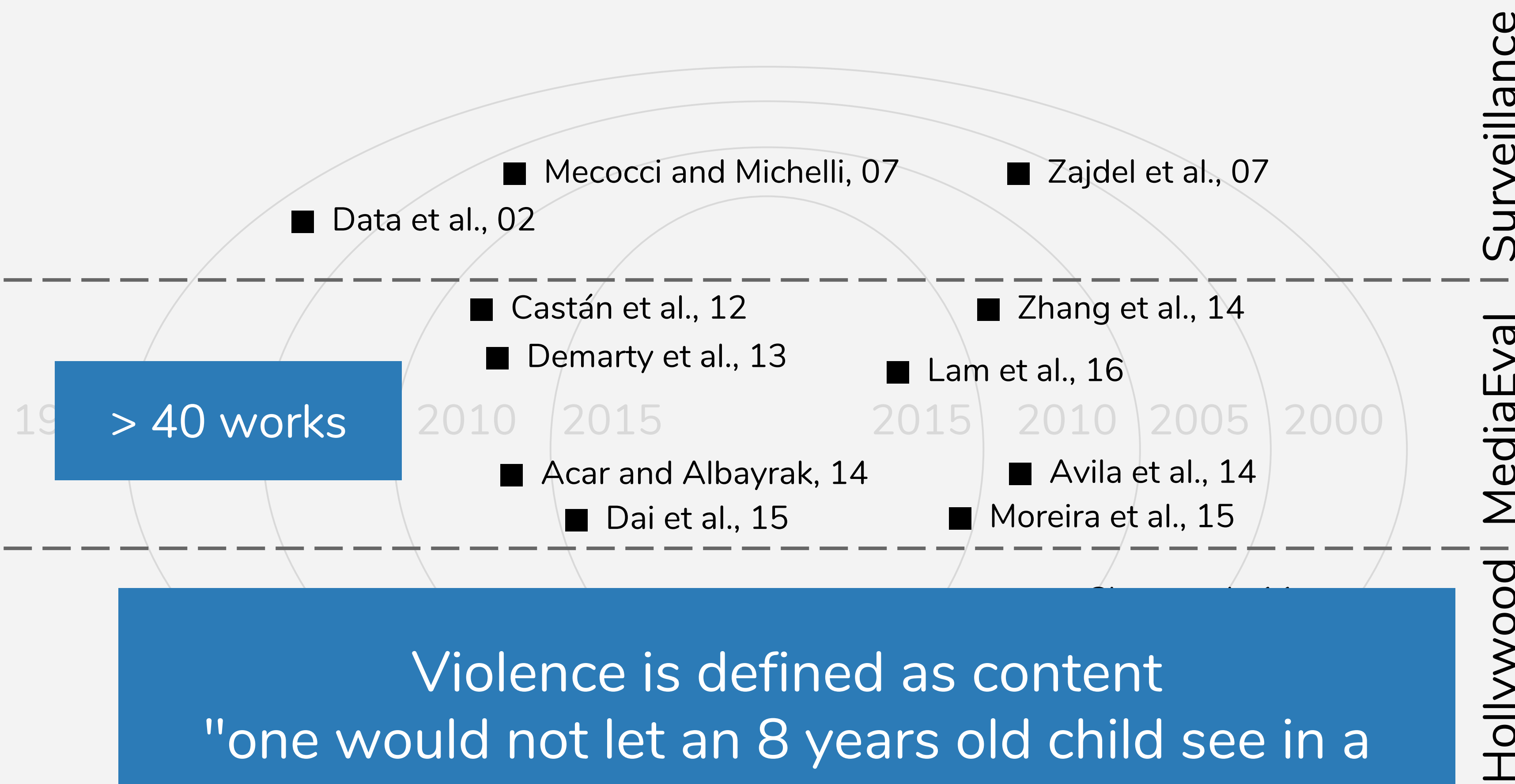
Violence Classification/Localization



Violence Classification/Localization



Violence Classification/Localization



> 40 works

Violence is defined as content
"one would not let an 8 years old child see in a video because it contains physical violence".

@Daniel Moreira 2016

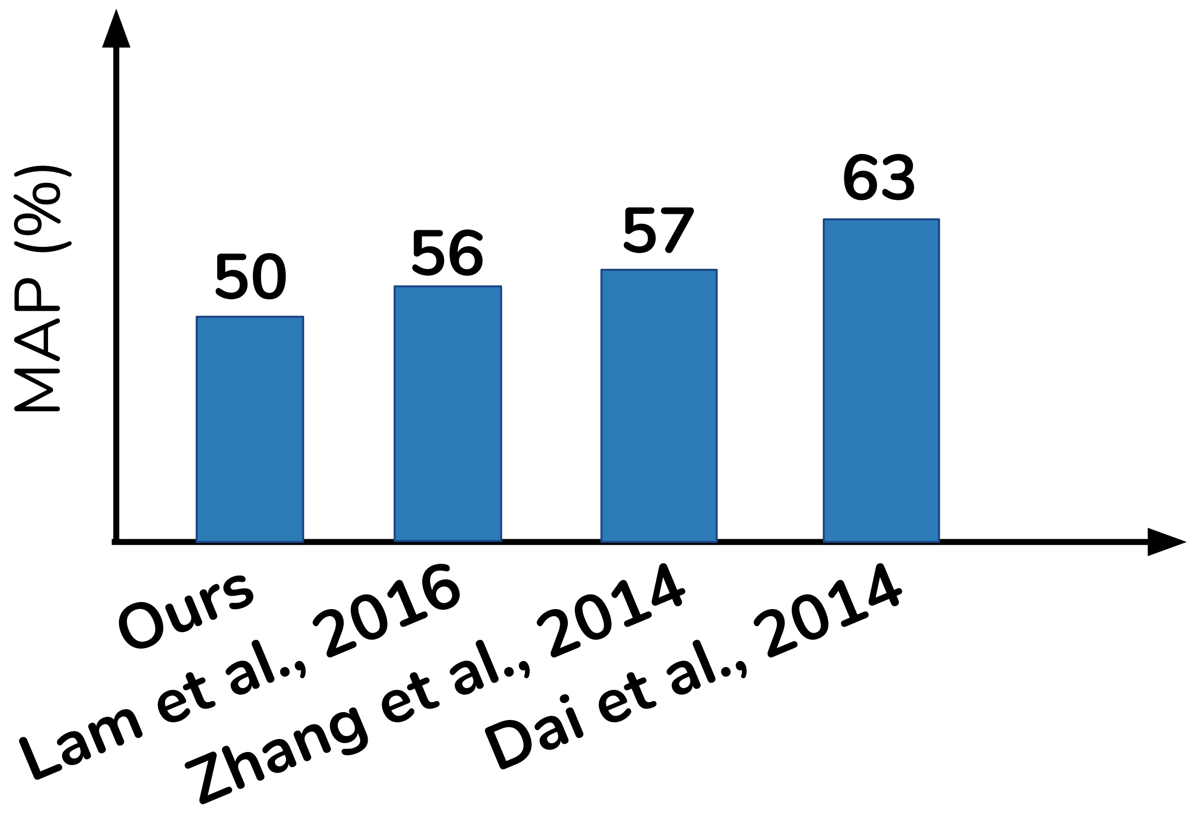
Dermaty et al. Benchmarking violent scenes detection in movies. In: IEEE CBMI, 2014.

Violence Classification/Localization

- **MediaEval 2011**
  - 50% ⇒ Bag-of-Visual-Words methods
- **MediaEval 2012**
  - 65% ⇒ Bag-of-Visual-Words methods
- **MediaEval 2013**
  - 90% ⇒ Bag-of-Visual-Words-based methods
- **MediaEval 2014**
  - 25% ⇒ Deep Learning-based methods
- **MediaEval 2015**
  - 80% ⇒ Deep Learning-based methods

Violence Localization

31 Hollywood movies
60 hours



Attack Approaches

Sensitive Media Analysis

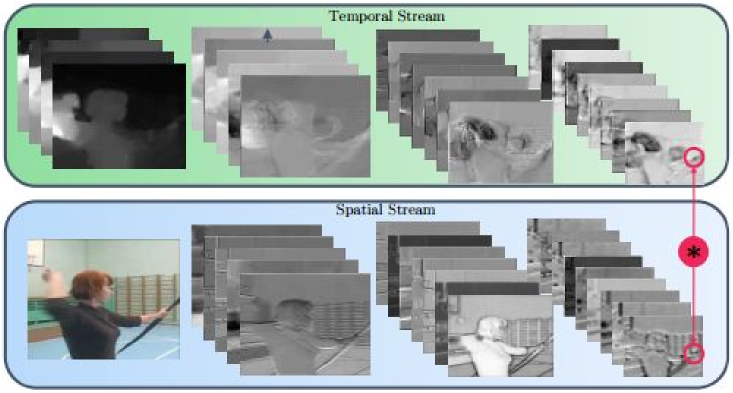
- **Temporal Deep Networks**
- **Deep Networks & Fisher Vectors**

Temporal Deep Networks

Challenge #1 — Pornography/Violence Localization

How to adapt Deep Networks for extracting space-temporal information?

- Few works have been investigating how to adapt Deep Networks for extracting space-temporal information

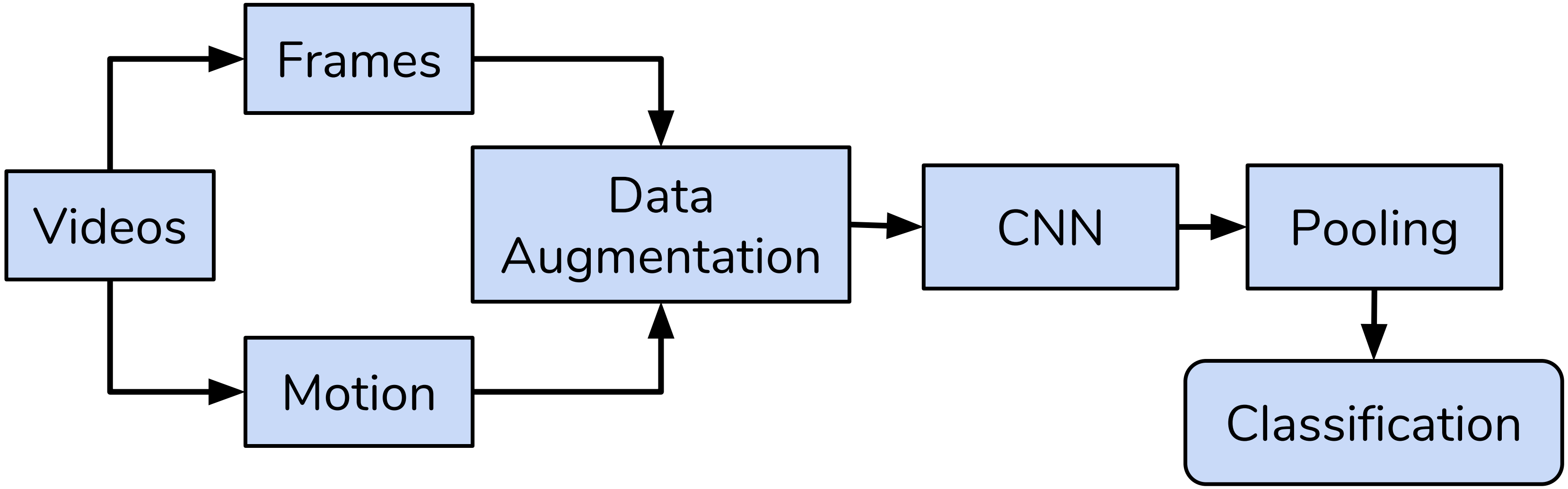


References
C. Feichtenhofer, A. Pinz, and A. Zisserman. Convolutional two-stream network fusion for video action recognition. In: CVPR, 2016.
Z. Liu, C. Zhang, and Y. Tian. 3D-based deep convolutional neural network for action recognition with depth sequences. Image and Vision Computing, 2016.

Temporal Deep Networks

Challenge #1 — Pornography/Violence Localization

- Explore data augmentation, data normalization, and dimensionality reduction
- Investigate different fusion methods (early, middle, late fusion)



Deep Networks & Fisher Vectors

Challenge #2 — Child Pornography Classification

Deep Networks ⇒ high accuracy

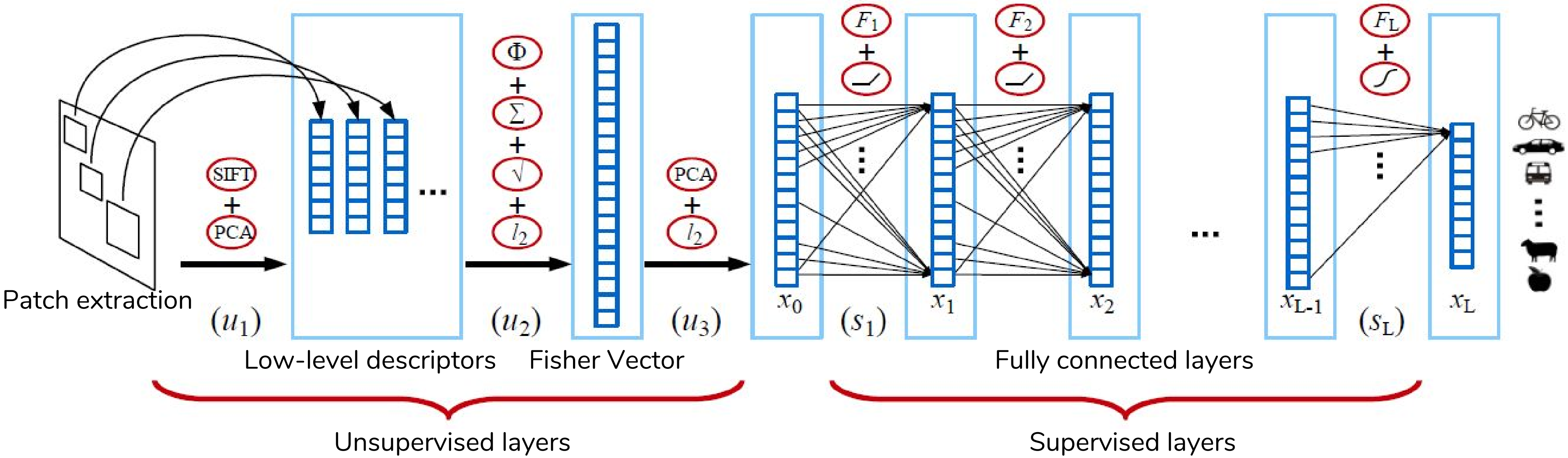
Fisher Vectors ⇒ less costly to train and evaluate

How to combine their strengths?

Deep Networks & Fisher Vectors

Challenge #2 — Child Pornography Classification

[Perronnin and Larlus, 2015]



References
F. Perronnin and D. Larlus. Fisher vectors meet neural networks: A hybrid classification architecture. In: CVPR, 2015.

Deep Networks & Fisher Vectors

Scientific Contributions

SC#1: Supervised dimensionality reduction

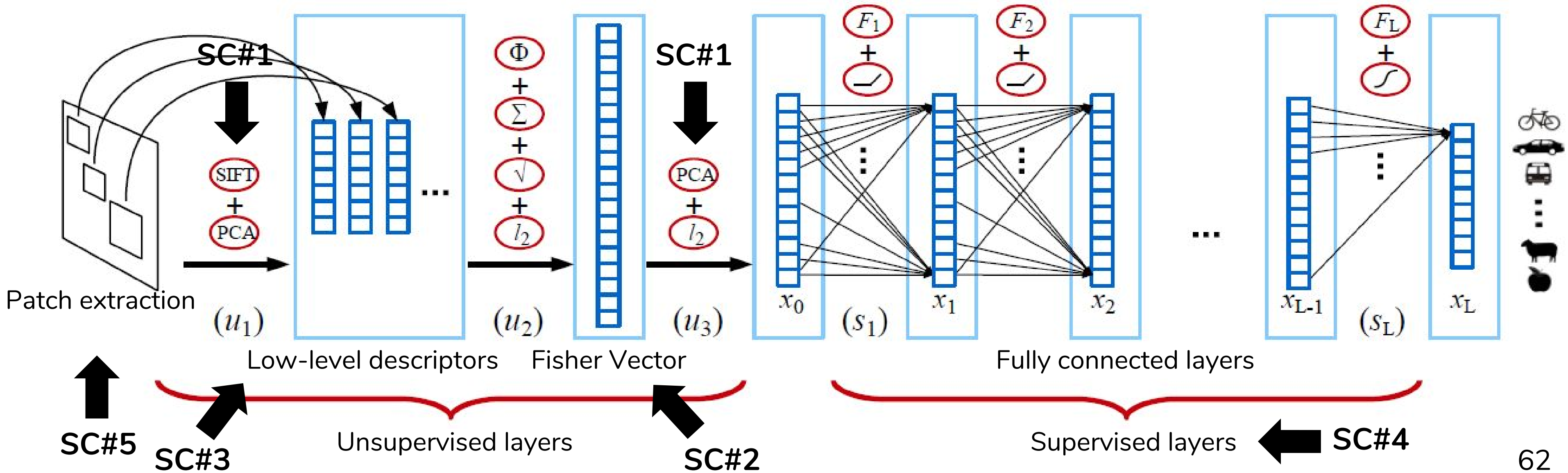
SC#2: Exponential family Fisher Vector

SC#3: Efficient local binary descriptors

SC#4: Exploring the network structural parameters

SC#5: Video classification

[Perronnin and Larlus, 2015]

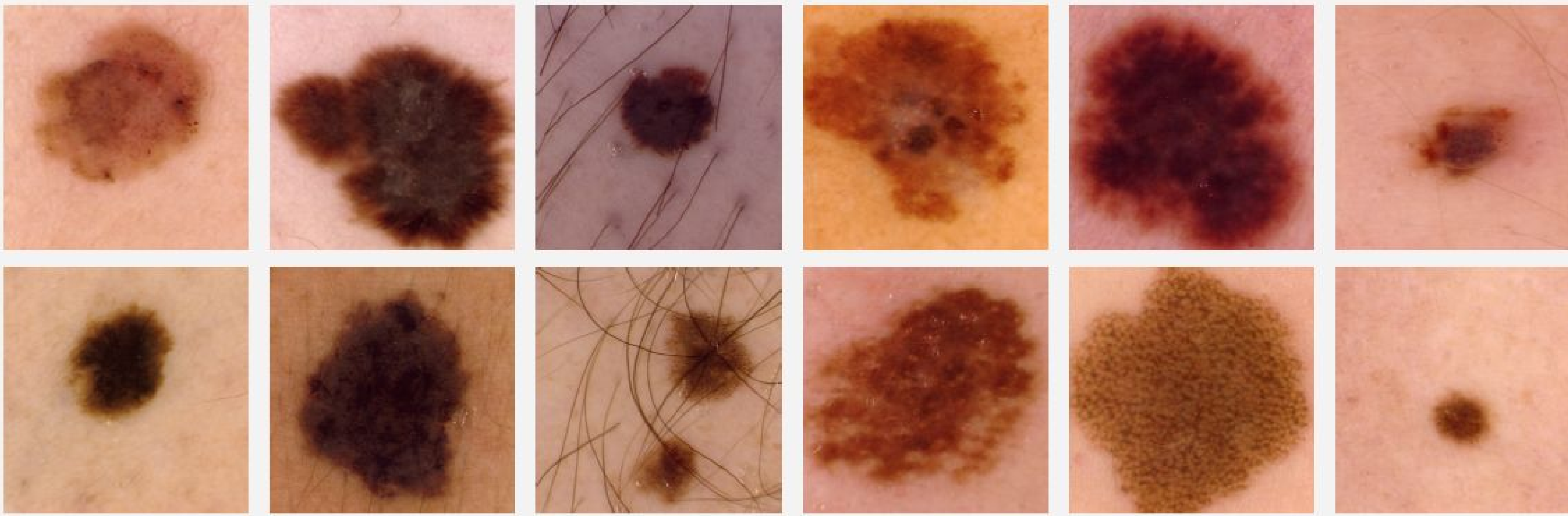


Our Applications
Melanoma Screening

"Melanoma is the type of **skin cancer** that most leads to **death**, but also the most **curable** if detected early."

American Cancer Society, 2017

Melanomas (top row) and
benign skin lesions (bottom row)



IRMA Dataset (747 images)

ISBI Challenge / ISIC Skin Lesion Analysis towards Melanoma Detection

ISIC 2016 Dataset (1,279 images)
(ISIC: International Skin Imaging Collaboration)

Literature
(Melanoma Screening)

Before ISBI Challenge (< 2016)

Ad-hoc image processing steps

Lesion segmentation
Hair removal
Border detection

Global features

Shape, color and texture features

Bag of Visual Words

Classical BoVW

After ISBI Challenge (>= 2016)

Deep learning-based methods

Premaladha and Ravichandran, 2016
Codela et al., 2016
Menegola et al., 2017
Berseth, 2017
Murphree et al., 2017
Zhang et al., 2017
Esteva et al, 2017

ISBI Challenge 2017

Equipe da Unicamp fica no topo de competição internacional de detecção automática de melanoma

| **Autor** Divulgação laboratório RECOD | **Fotos** Mijail Vidal | **Edição de imagem** Paulo Cavalheri

Uma equipe de professores e pesquisadores da Unicamp obteve excelente resultado na segunda edição da Competição Internacional de Análise de Lesões de Pele, evento anual não-presencial organizado pela Colaboração Internacional para Imagens de

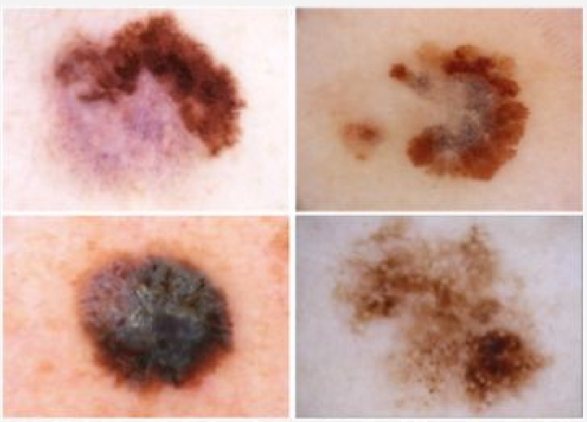
Our Solution

(Melanoma Screening)

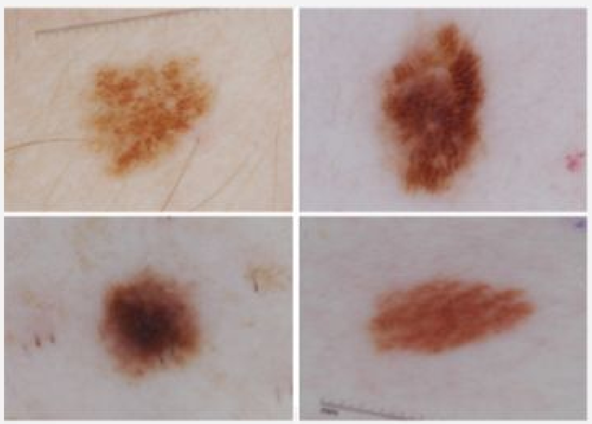
ISBI Challenge 2017

Lesion Classification

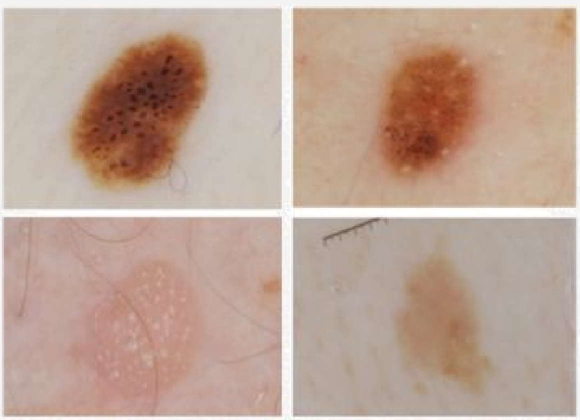
Melanoma



Nevus



Seborrheic keratosis



(1) Melanoma vs. Nevus and Seborrheic keratosis
(2) Seborrheic keratosis vs. Melanoma and Nevus

ISBI Challenge 2017
Lesion Classification

- Our previous experience with the technique taught us that **three big bottlenecks** would limit performance:
  - Amount of training data,
  - Depth of the learning model, and
  - Availability of computational horsepower.

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

ISBI Challenge 2017
Lesion Classification

- Amount of **training data**
  - ISIC 2017 Challenge: **2,000** dermoscopic images
  - ISIC Archive: **13,000** dermoscopic images
  - Interactive Atlas of Dermoscopy: **1,000+** clinical cases
  - Dermofit Image Library: **1,300** images
  - IRMA Skin Lesion Dataset: **747** dermoscopic images
  - PH2 Dataset: **200** dermoscopic images

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

ISBI Challenge 2017
Lesion Classification

- An initial agenda of hypotheses to validate.
  1. Compare the baseline **VGG-16** network to the deeper **ResNet-101** or **Inception-v4**;
  2. Compare standard-resolution images to double-resolution images;
  3. Contrast different strategies of class- and sample-weighting during training;
  4. Compare normal training schedule with some form of curriculum-learning;

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

ISBI Challenge 2017
Lesion Classification

- An initial agenda of hypotheses to validate.
  5. Contrast different regimens of training and test augmentation;
  6. Measure the impact of adding SVM as a final decision layer;
  7. Attempt to use the patient data (age and sex) on classification;
  8. Attempt different model optimizers;
  9. Add different types of per-sample normalization;
  10. Add a final meta-decision based upon multiple models (ensemble, stacking, etc.)

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

ISBI Challenge 2017
Lesion Classification

The biggest **disappointments/surprises**:

1. Image resolution
2. Sample-weighting schemes
3. Validation and early stopping
4. Patient data
5. Curriculum-learning
6. Segmentation information

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

ISBI Challenge 2017
Lesion Classification

Success factors

1. Models + data
2. Data augmentation
3. Per-image normalization
4. Stacking models and meta-learning

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

ISBI Challenge 2017

Lesion Classification

Rank	Organization	Score
1	RECOD Titans / Unicamp	0.874
2	USYD-BMIT	0.870
3	Casio and Shinshu University	0.868
4	Universidad Carlos III de Madrid	0.856
5	University of Guelph - MLRG	0.836

A. Menegola, J. Tavares, M. Fornaciali, L. Li, S. Avila, E. Valle. RECOD Titans at ISIC Challenge 2017. arXiv, 2017.

Attack Approaches
Melanoma Screening

ISBI Challenge 2017
Lesion Classification

The biggest **disappointments/surprises**:

1. Image resolution
2. Sample-weighting schemes
3. Validation and early stopping
4. Patient data
5. Curriculum-learning
6. Segmentation information

Final Remarks

Final Remarks

- Hot topics in Machine Learning research
- Good perspectives for future contributions
- Social impact research

Sandra Avila
www.ic.unicamp.br/~sandra