IM●GENET
LSVRC Winners
2010 - 2015

Presented by Weilin

IM🟢GENET LSVRC

- **ImageNet**: an image database organized according to the **WordNet** hierarchy (currently only the nouns), ~500 images per node.

- **LSVRC**: Large Scale Visual Recognition Challenge based on ImageNet. Often referred as *ImageNet Competition/Challenge*.

Common Tasks

Common Tasks

Image Classification

2010   2011   2012   2013   2014   2015   2016   ......

Object Localization

Bounding box prediction.
Require >50% overlap.

Common Tasks

Image Classification

AlexNet started the Deep Learning era.

2010　　2011　　2012　　2013　　2014　　2015　　2016　　……

Object Localization

Common Tasks

Image Classification

AlexNet started the Deep Learning era.

2010   **2011**   **2012**   **2013**   **2014**   **2015**   **2016**   ......

Object Localization

Object Detection
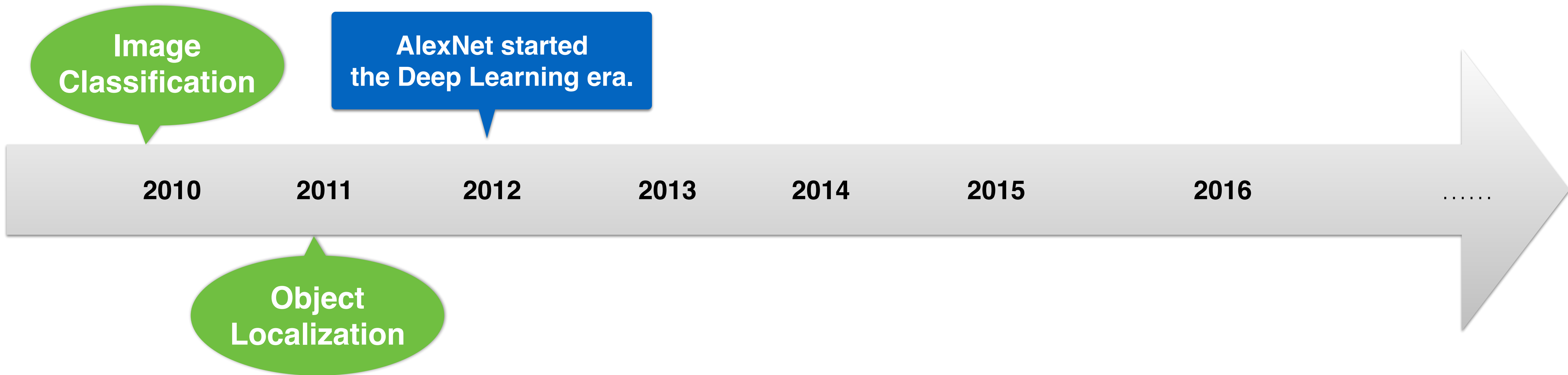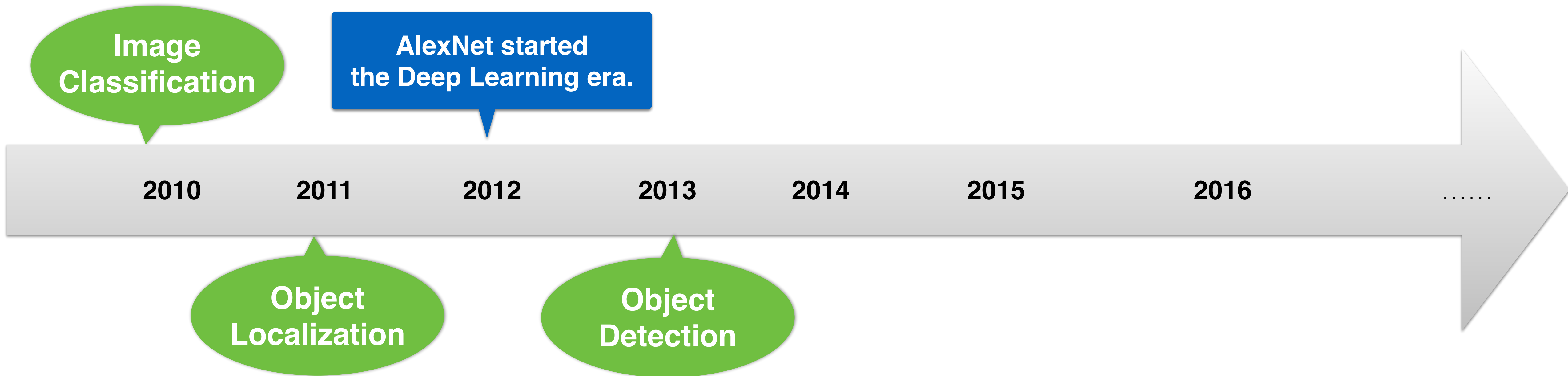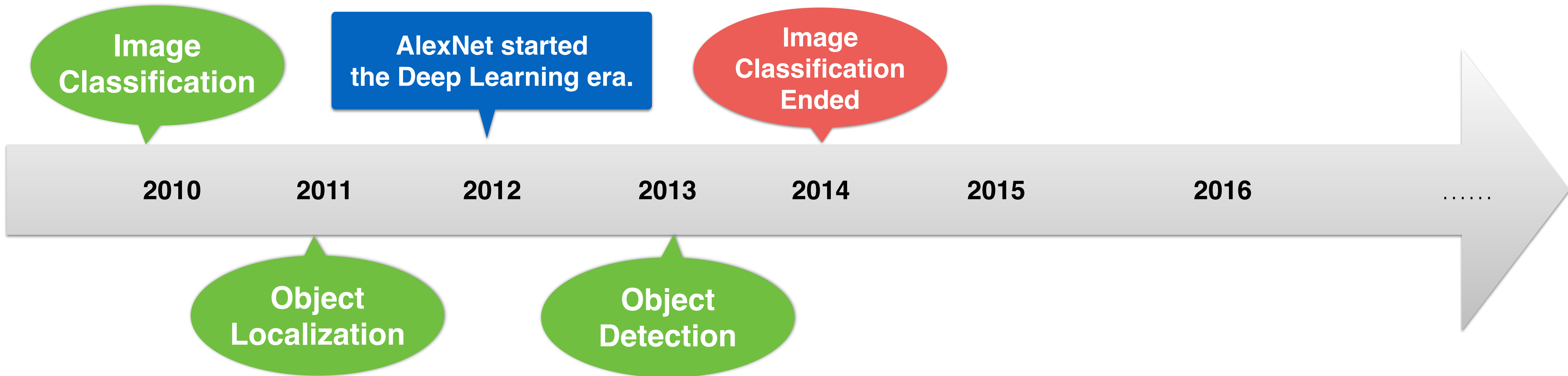
Different metrics: Penalize false positives.
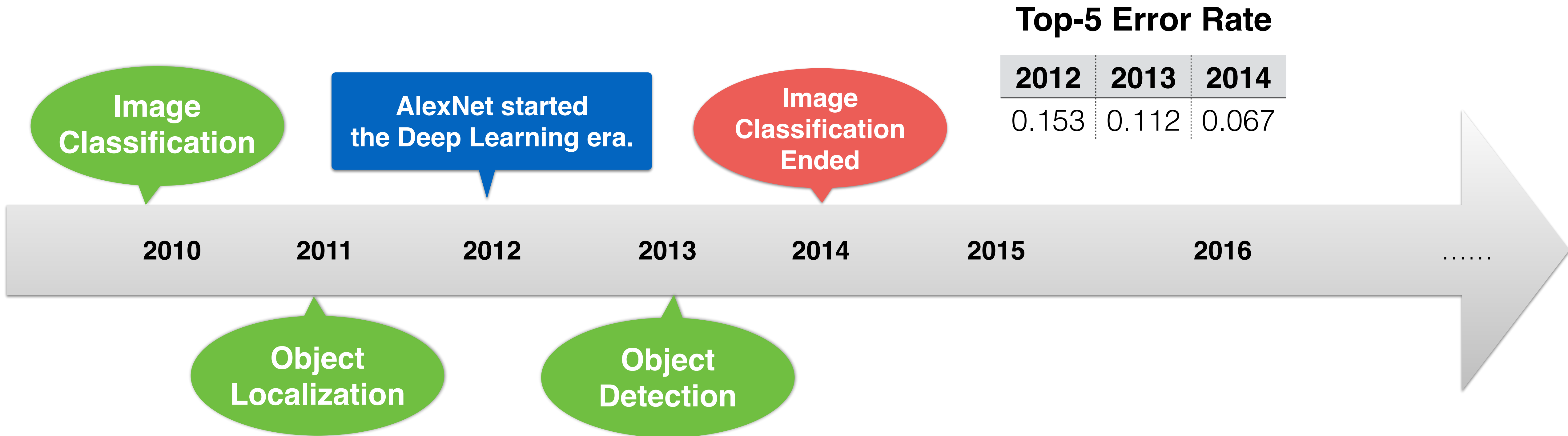Thus, need to detect background.
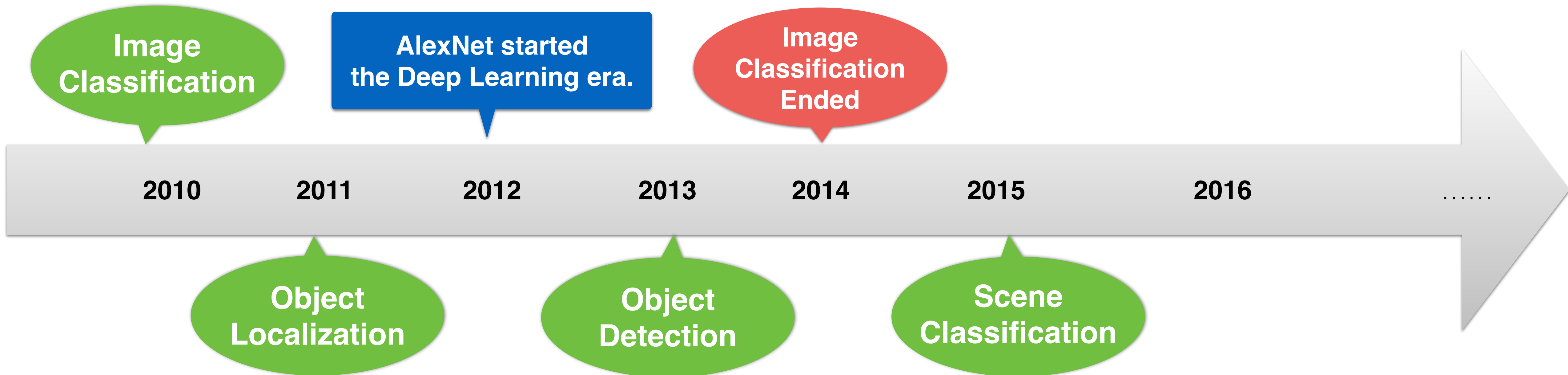Fewer classes (200), but more (or none) objects and smaller objects.

Common Tasks

Common Tasks

Top-5 Error Rate

2012	2013	2014
0.153	0.112	0.067

Image Classification — 2010

AlexNet started the Deep Learning era. — 2012

Image Classification Ended — 2014

Object Localization — 2011

Object Detection — 2013
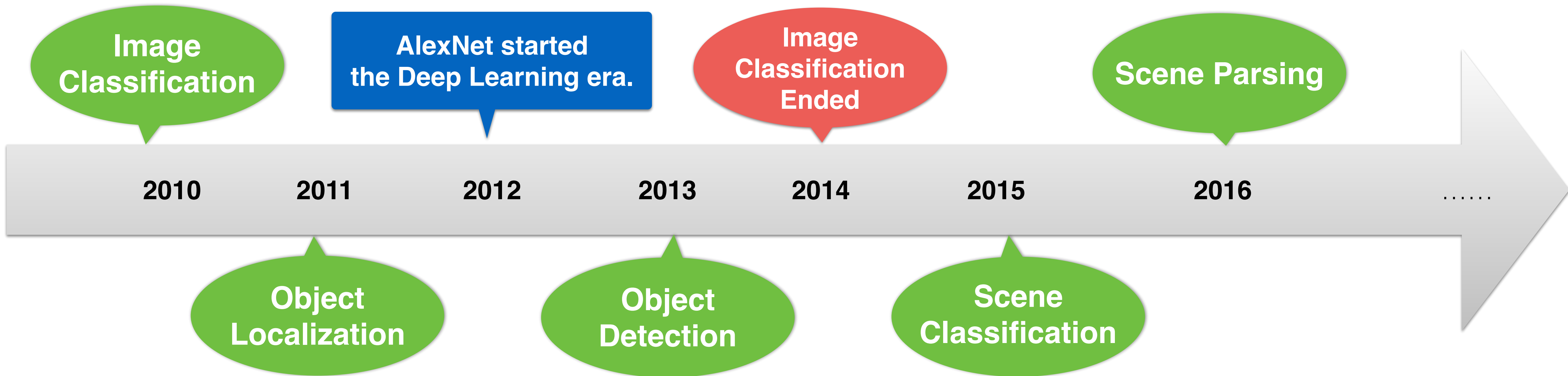
2010 2011 2012 2013 2014 2015 2016 ......
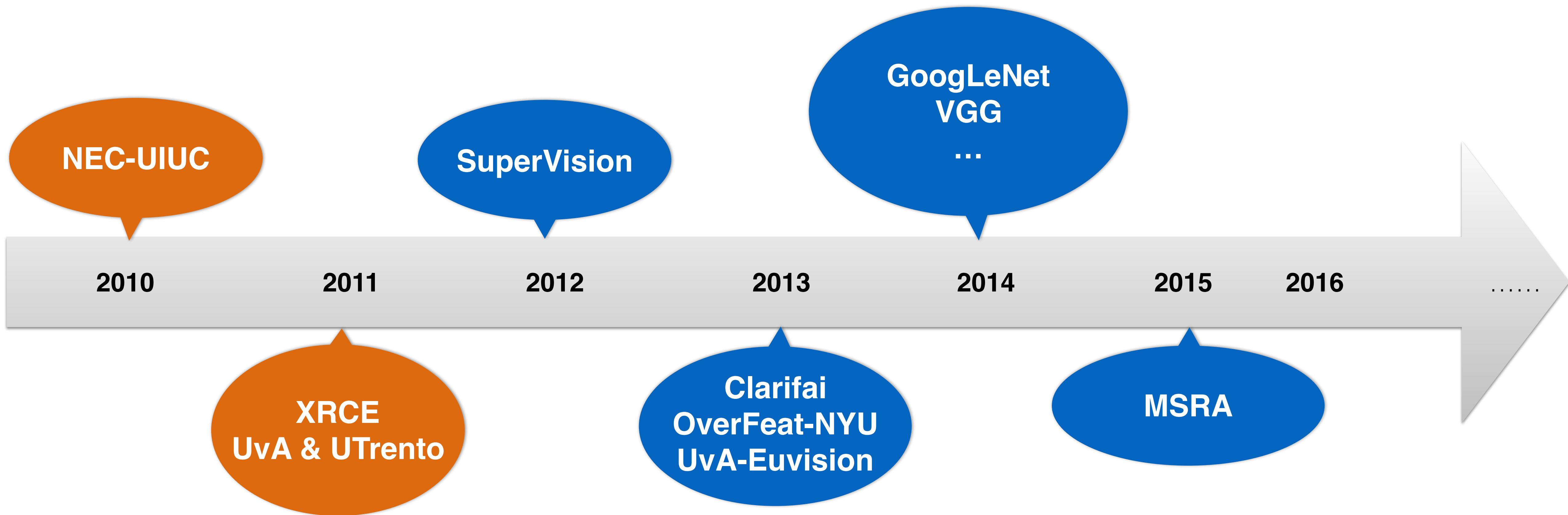
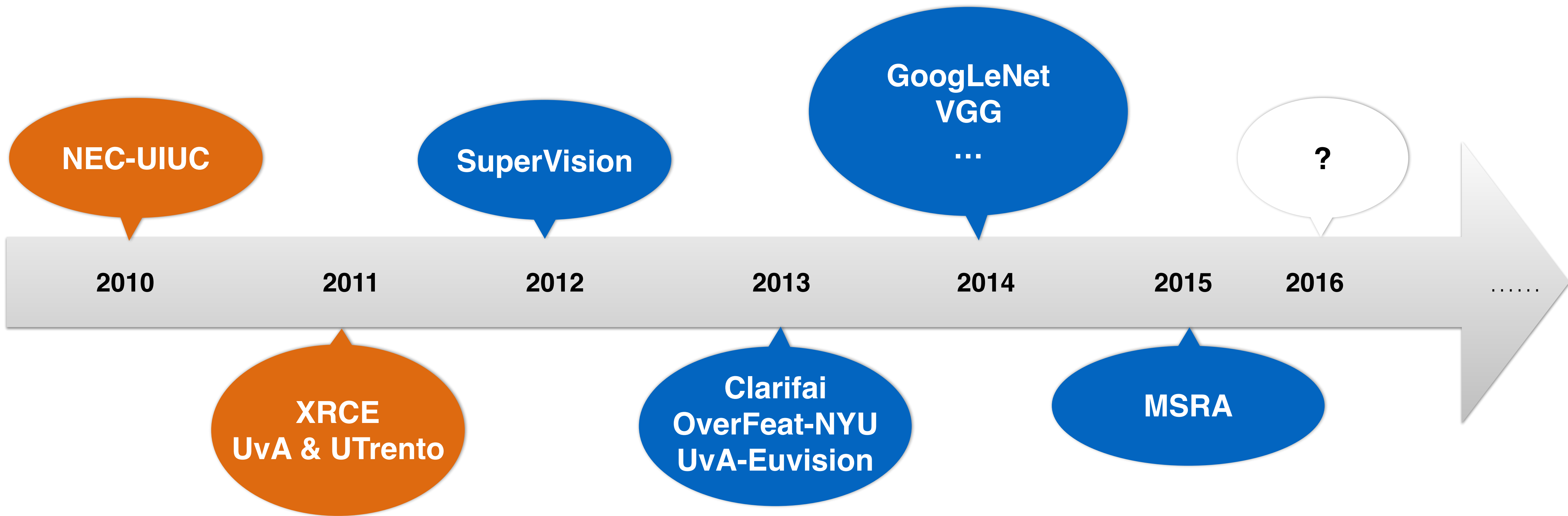# Common Tasks

Common Tasks

Dataset Scale

- **Localization Dataset** (unchanged since 2012)

  - Classes: 1000 categories

  - Training: 1,200,000 labeled images (~1200 images per class)

  - Validation: 50,000 labeled images (~50 images per class)

  - Evaluation: 100,000 images without labels (~100 images per class)

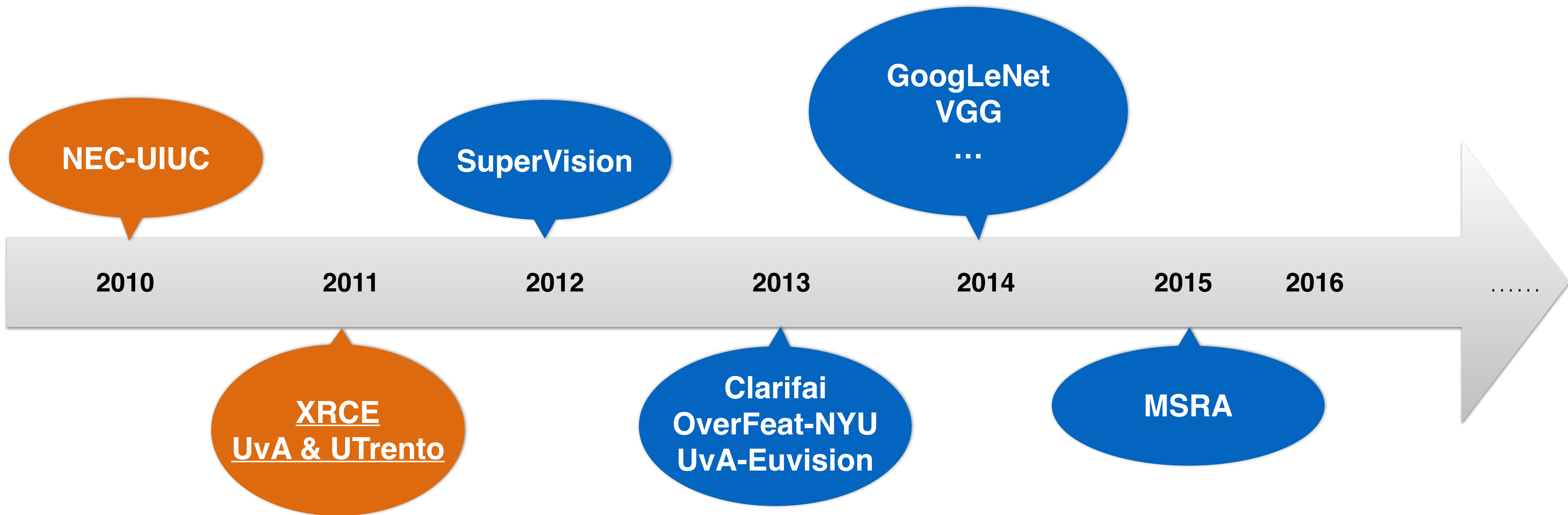# Winners since 2010

Winners since 2010

2010: NEC-UIUC

- Yuanqing Lin, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu (NEC).

- LiangLiang Cao, Zhen Li, Min-Hsuan Tsai, Xi Zhou, Thomas Huang (UIUC).

- Tong Zhang (Rutgers).

- Link to Slides

2010: NEC-UIUC

- **Low-level Features**: LBP + HOG

  - Local Binary Patterns, Histogram of Oriented Gradients

- **Sparse Coding**: Local coordinate coding (LCC) / Super-vector coding

- **Linear SVM** Classification with **ASGD**

  - Feature dimension: 82K to 262K   Images: 1.2M

  - Classes: 1000

Winners since 2010

NEC-UIUC

SuperVision

GoogLeNet
VGG
…

2010 **2011** **2012** **2013** **2014** **2015** **2016**

XRCE
UvA & UTrento

Clarifai
OverFeat-NYU
UvA-Euvision

MSRA

2011 Classification: XRCE

- Florent Perronnin, Xerox Research Centre Europe (XRCE)

- Jorge Sanchez, XRCE / CIII

- [Link to Slides](#)

['zɪ.rɑks]

2011 Classification: XRCE

- **Low-level Features**: 128-d SIFT & 96-d Color, reduced to 64-d with PCA

- Image signatures with **Fisher Vectors** (FV)

    - Output: high-dimensional (e.g. 0.5M dim) and weakly sparse (e.g. half non-zeros)

- Two possible solutions:

    - Dimensionality reduction with PCA, etc. Accuracy Dropped :(

    - Coding with **Product Quantization** (PQ), reduce storage only :)

- **Linear SVM** in the space of FVs.

2011 Localization: UvA & UTrento

- Koen E. A. van de Sande, University of Amsterdam

- Jasper R. R. Uijlings, University of Trento

- Arnold W. M. Smeulders, University of Amsterdam

- Theo Gevers, University of Amsterdam

- Nicu Sebe, University of Trento

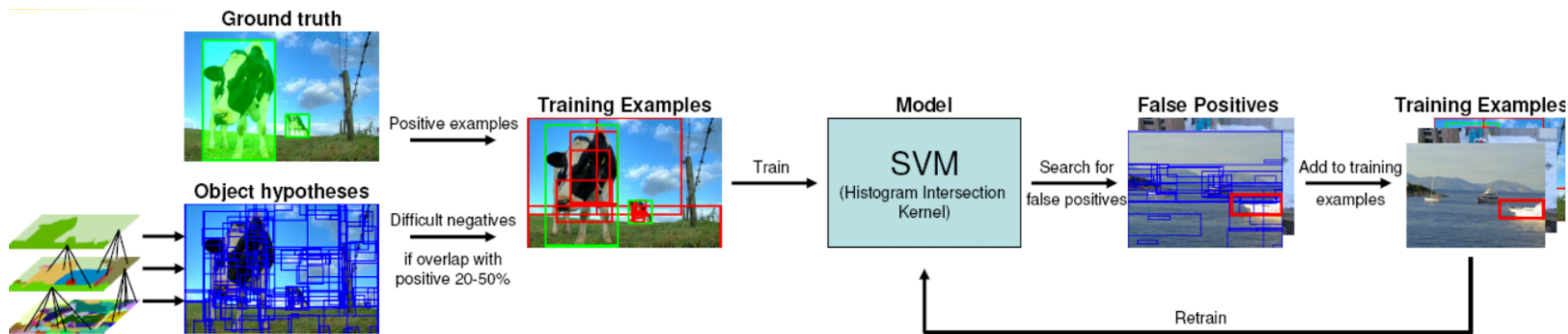- Cees Snoek, University of Amsterdam

- Link to Slides
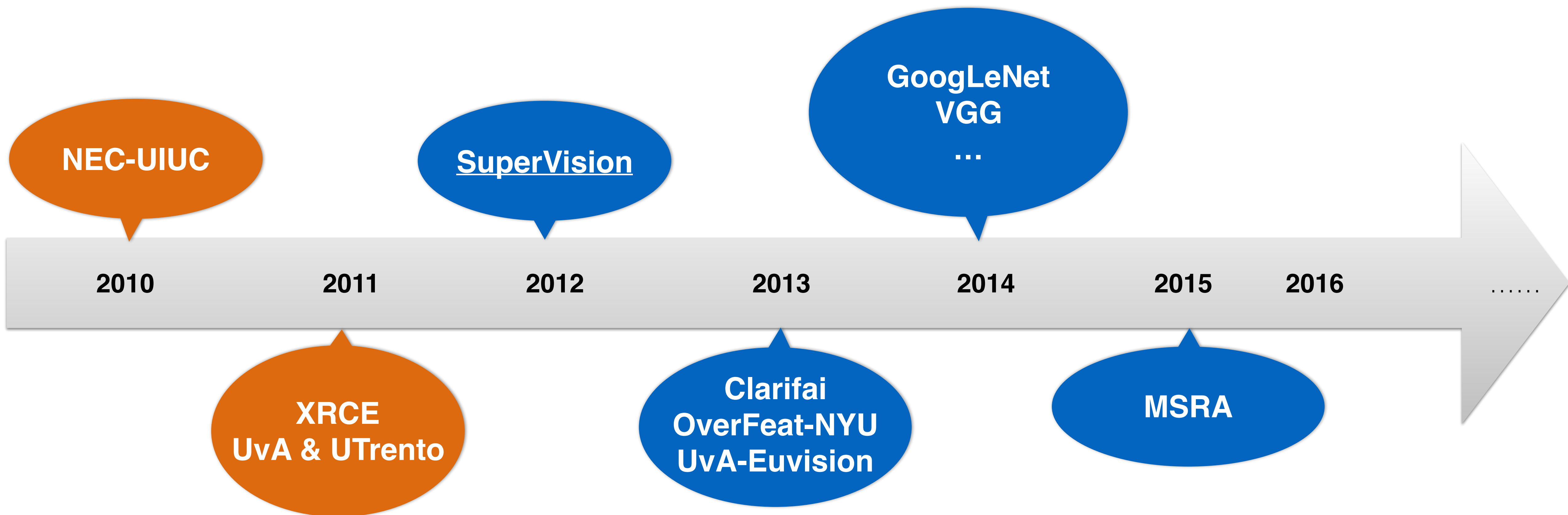
2011 Localization: UvA & UTrento

- **Selective Search** (vs. Exhaustive Search),

  - Emphasized high recall, efficiency (<10s per image)

  - (Also used in R-CNN for region proposals)

Winners since 2010

NEC-UIUC

XRCE
UvA & UTrento

SuperVision

Clarifai
OverFeat-NYU
UvA-Euvision

GoogLeNet
VGG
...

MSRA

2010   2011   2012   2013   2014   2015   2016   ......
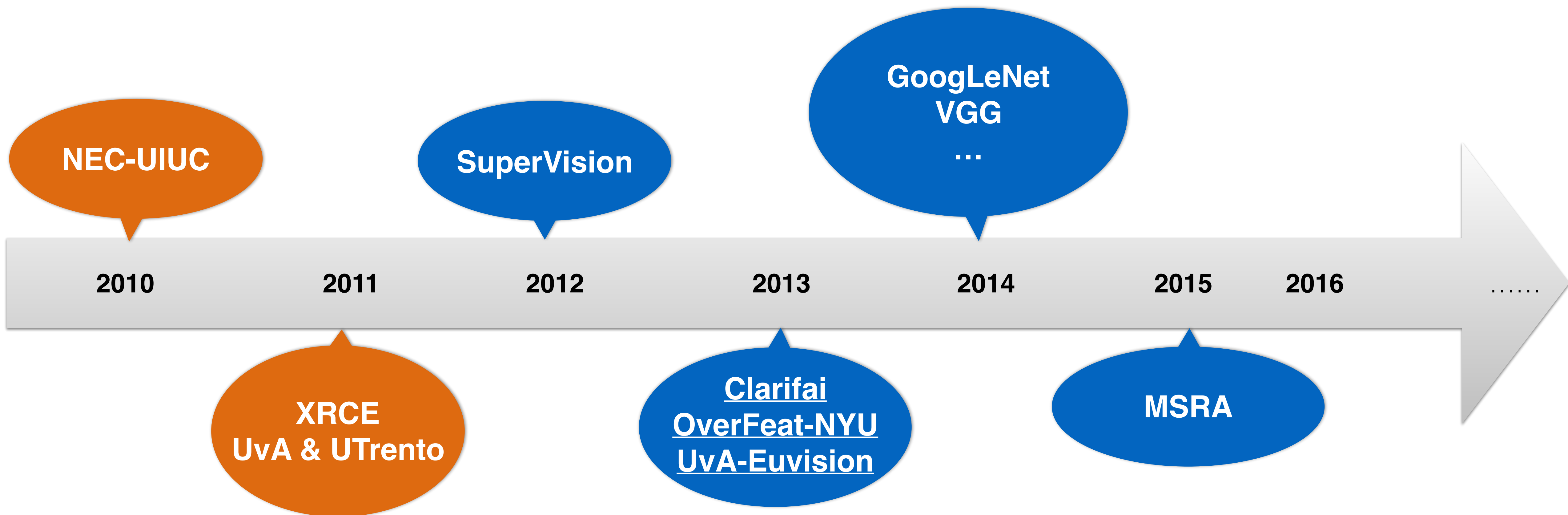
2012 Winner: SuperVision (AlexNet)

- Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton

- University of Toronto

- Top-5 error: 0.15 Vs. 0.26 (2nd place)

2012 Winner: SuperVision (AlexNet)

- 5-layer Convolution + 3-layer FC

- Overlapping Pooling

- ReLU Nonlinearity (often tanh before)

- Local Response Normalization ("brightness normalization")

  - Batch Normalization was in 2015

- Dropout

- Multi-GPU Training

Winners since 2010

NEC-UIUC

XRCE
UvA & UTrento

SuperVision

Clarifai
OverFeat-NYU
UvA-Euvision

GoogLeNet
VGG
…

MSRA

2010   2011   2012   2013   2014   2015   2016   ……

2013 Winners

- Classification: **Clarifai**

- Localization: **OverFeat-NYU**

- Detection: **UvA-Euvision**

2013 Classification: Clarifai

- Preprocessing: subtracting a per-pixel mean (per-image?)

- Data augmentation:

  - Downsample to 256 pixel

  - Random 224 pixel crops

  - Random horizontal flipping

- Dropout.

- Model selection with a novel visualization technique based on the deconvolutional networks
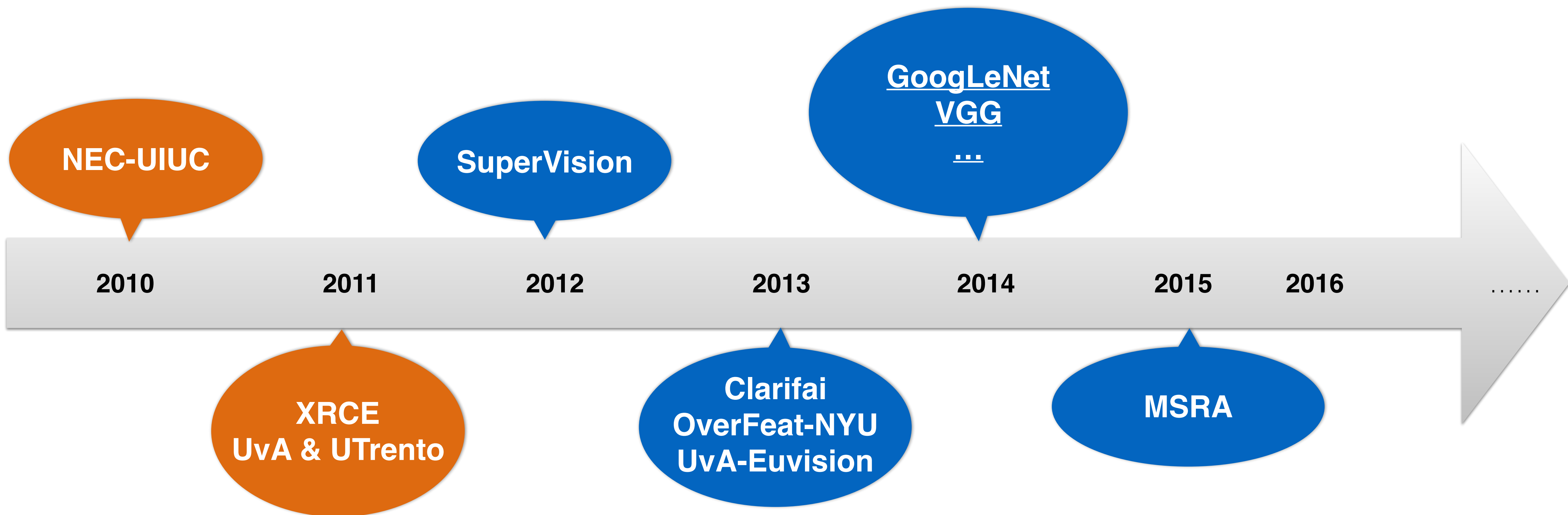
2013 Localization: OverFeat-NYU

- A little different from AlexNet:

  - No Contrast Normalization (No Local Norm)

  - Non-overlapping Pooling

  - Smaller stride (4->2), thus larger feature maps in 1st & 2nd layer

- Multi-Scale Classification (?)

- Details on CNN Localization first time

  - Integrated Learning with Classification.

  - Per-class bounding box regression

2013 Detection: UvA-Euvision

- **Selective Search** for region proposals.

- Boxes are represented by **SIFT** Features at multiple scales.

- An efficient **encoding** on boxes. (paper in submission at that time)

- **ConvNet** trained on DET dataset for object presence priors.

Winners since 2010

NEC-UIUC

XRCE
UvA & UTrento

SuperVision

Clarifai
OverFeat-NYU
UvA-Euvision

GoogLeNet
VGG
...

MSRA

2010 2011 2012 2013 2014 2015 2016 ......

2014 Winners

- Task 1a: Object detection**: NUS**

- Task 1b: Object detection with additional training data: **GoogLeNet**

- Task 2a: Localization: **VGG**

- Task 2b: Localization with additional training data: **Adobe-UIUC**

2014 Detection: NUS

- Jian DONG, Yunchao WEI, min LIN, Wei XIA, Shuicheng YAN (National University of Singapore), Qiang CHEN (IBM Research, Australia)

- **Network In Network** (NIN) as feature extractor. (?)

  - A special modification of CNN, in ICLR 2014

- Augmented training and testing sample.

- Integration of output from traditional framework with SVM.

- Detection based on **R-CNN**

2014 Detection: GoogLeNet

- 22-layer deep network: 21 Conv + 1 FC

- Based on Inception v1

- **Inception** module

  - Multi-scale idea

  - **Hebbian principle**: "Cells that fire together, wire together."

- Local Response Normalization; No Dropout.

- Detection like **R-CNN**.
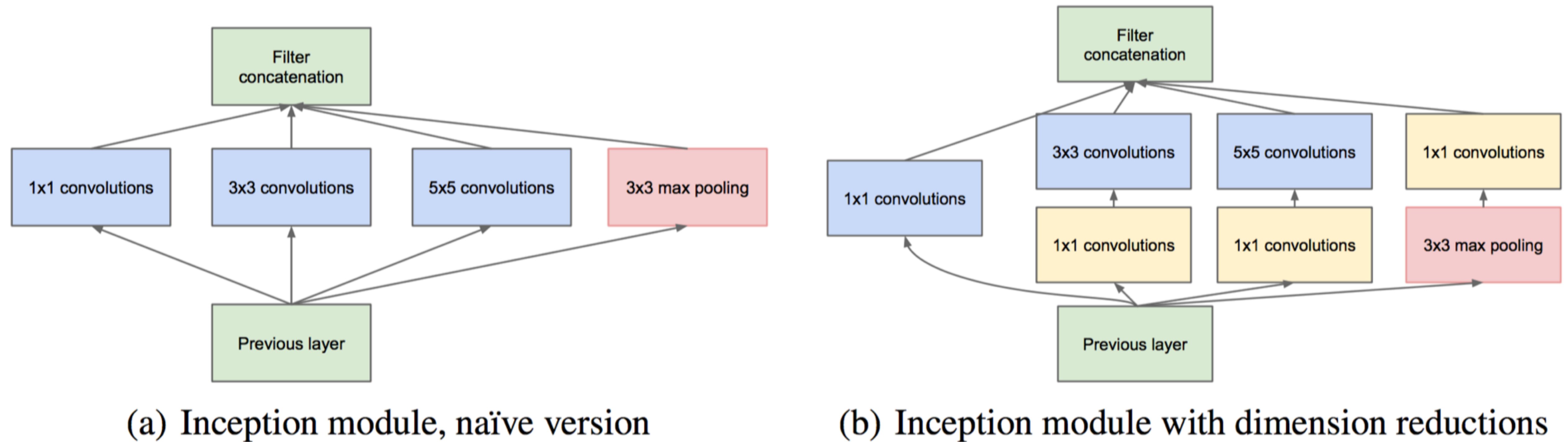
2014: GoogLeNet (Inception v1)

(a) Inception module, naïve version

(b) Inception module with dimension reductions
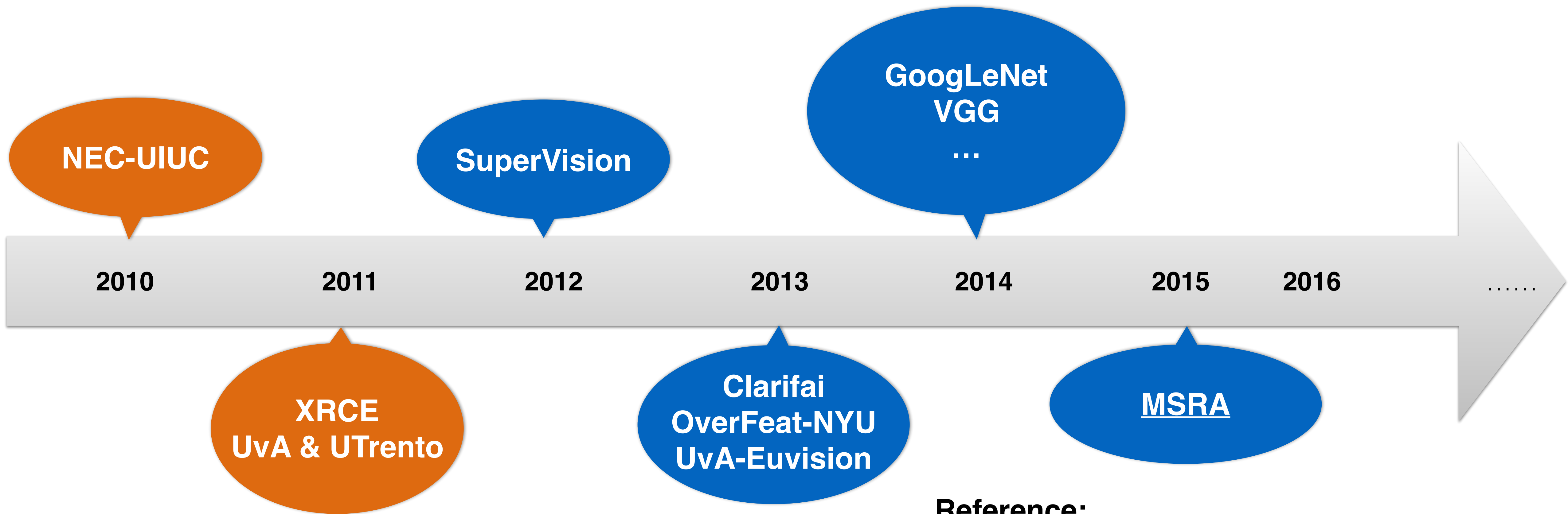
Figure 2: Inception module

2014 Localization: VGG (OxfordNet)

- Karen Simonyan, Andrew Zisserman (University of Oxford)

- Runner-up in 2013

- Nothing special on network architecture.

  - Tried hard in fine-tuning hyper parameters?

- Train with crops, test with uncropped images.

- Localization: per-class bounding box regression similar to OverFeat

2014 Localization: Adobe-UIUC

- Hailin Jin, Jianchao Yang, Zhe Lin (Adobe), Zhaowen Wang (UIUC)

- An integrated CNN for both classification and localization (like OverFeat)

- K-means to find bounding box clusters and rank by classification score.

Winners since 2010

NEC-UIUC

SuperVision

**GoogLeNet
VGG
…**

2010   2011   2012   2013   2014   2015   2016   ……

**XRCE
UvA & UTrento**

**Clarifai
OverFeat-NYU
UvA-Euvision**

MSRA

Reference:

ICML 2016 Tutorial on Deep Residual Training

36