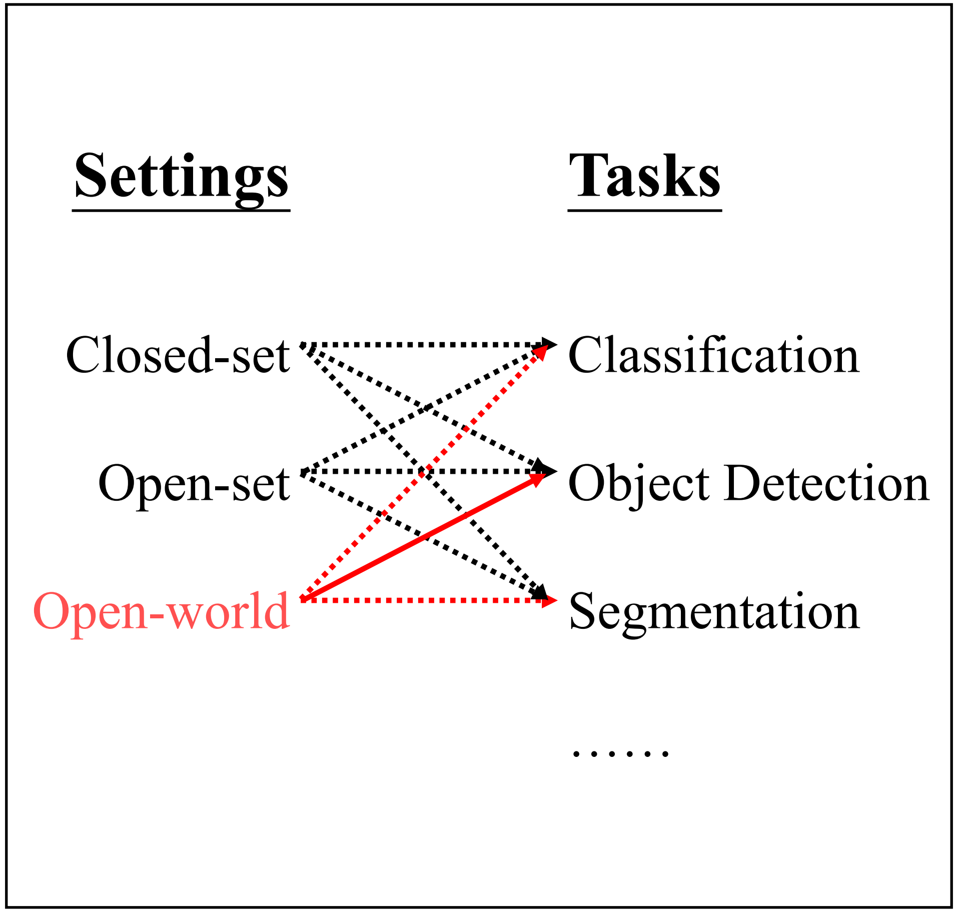
An Introduction to

Open World Object Detection

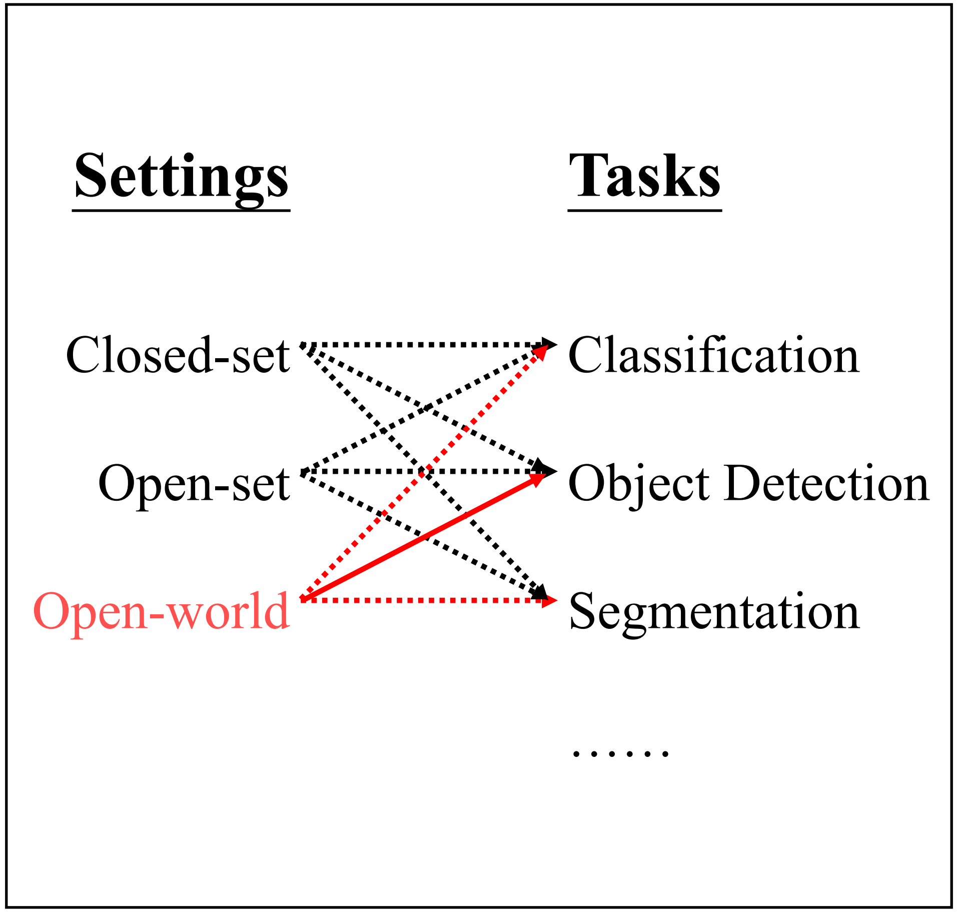
Presenter: Liu Dai
05/14/2022

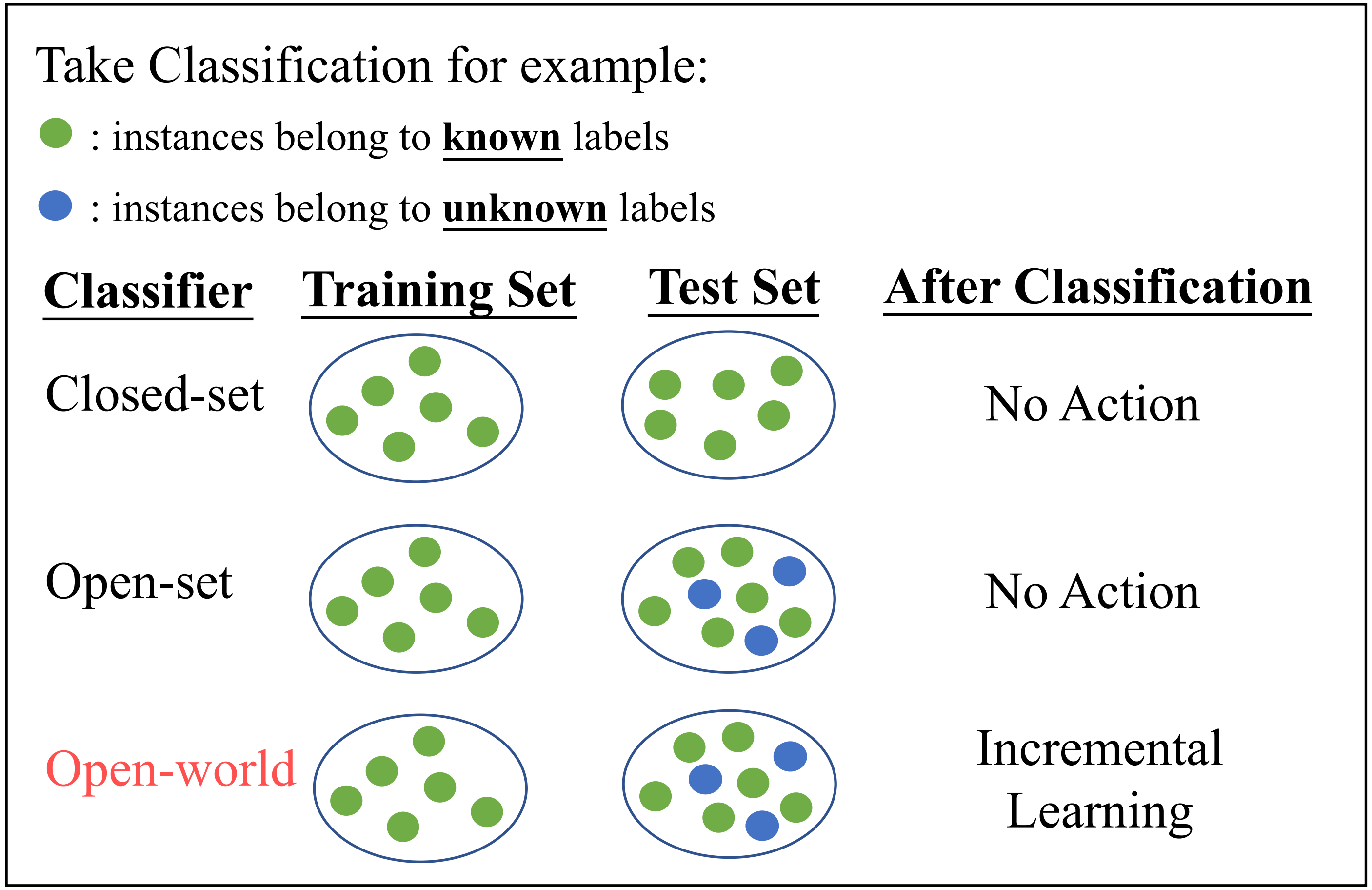
CONTENTS

1. Open-World Setting



1. Open-World Setting



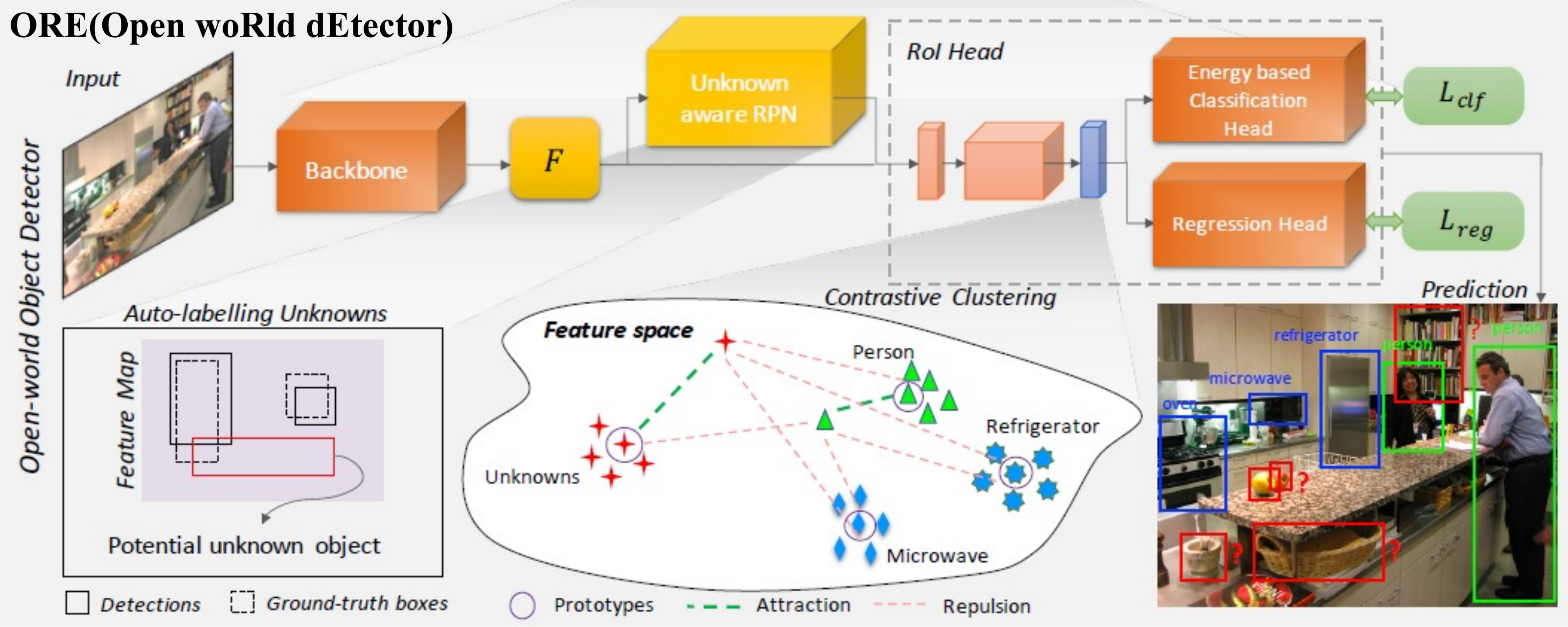


An open-world classifier across its life span:

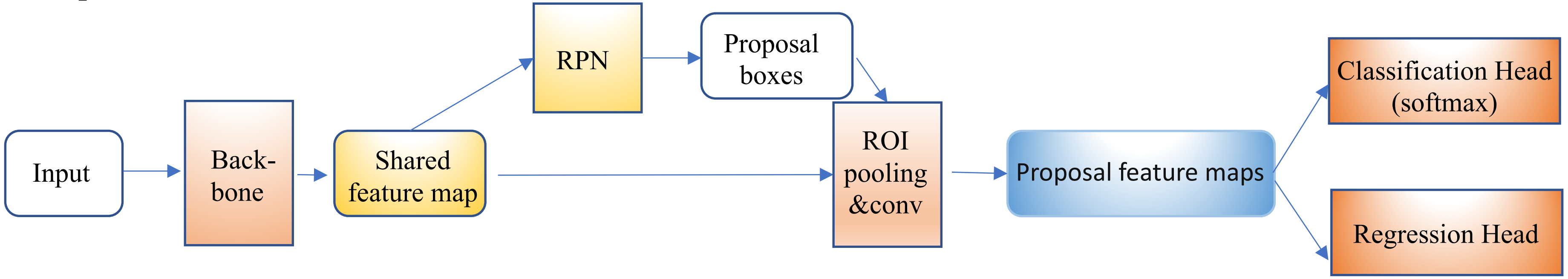
…… → Classify known classes and pick out unknowns → Oracle Labelling → Incremental Learning → Classify known classes and pick out unknowns → ……

2. Recap of *Towards Open World Object Detection*

******Towards Open World Object Detection*****,** KJ Joseph et al, IIT, CVPR2021 Oral **(Original OWOD)**



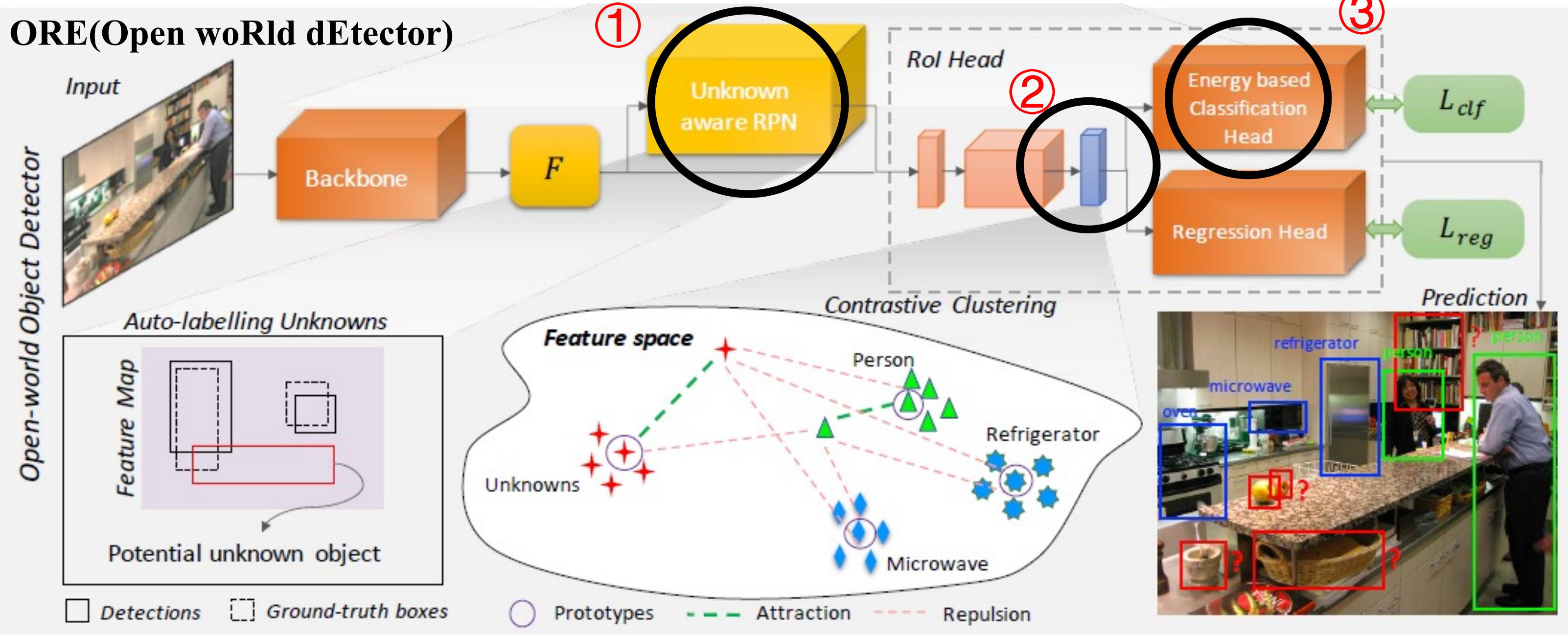
Compared with Classical Faster-RCNN:



(*) The article has been cited for **61** times until 05/14/2022.

2. Recap of *Towards Open World Object Detection*

Towards Open World Object Detection*, KJ Joseph et al, IIT, CVPR2021 Oral **(Original OWOD)



Based on Faster-RCNN,

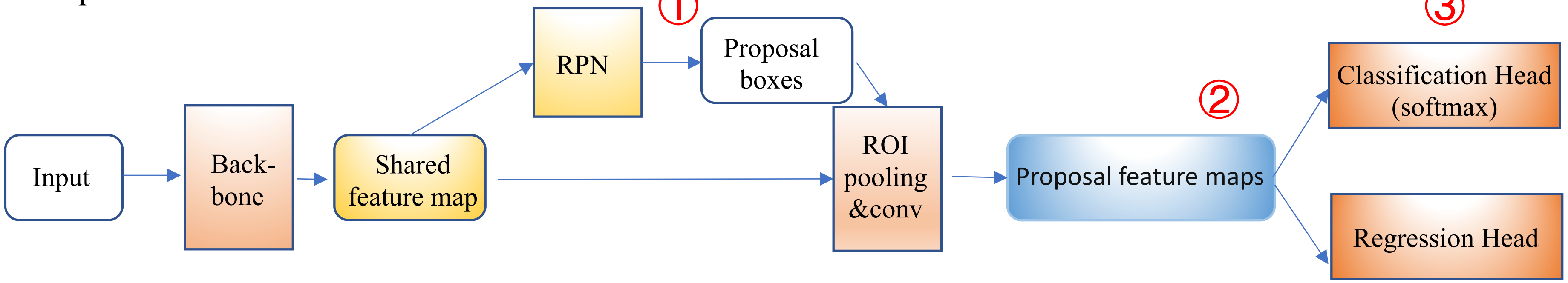
ORE adds 3 new modules:

① Auto-labelling unknown regions with RPN

② Contrastive Clustering

③ Energy-based Unknown Identifier(EBUI)

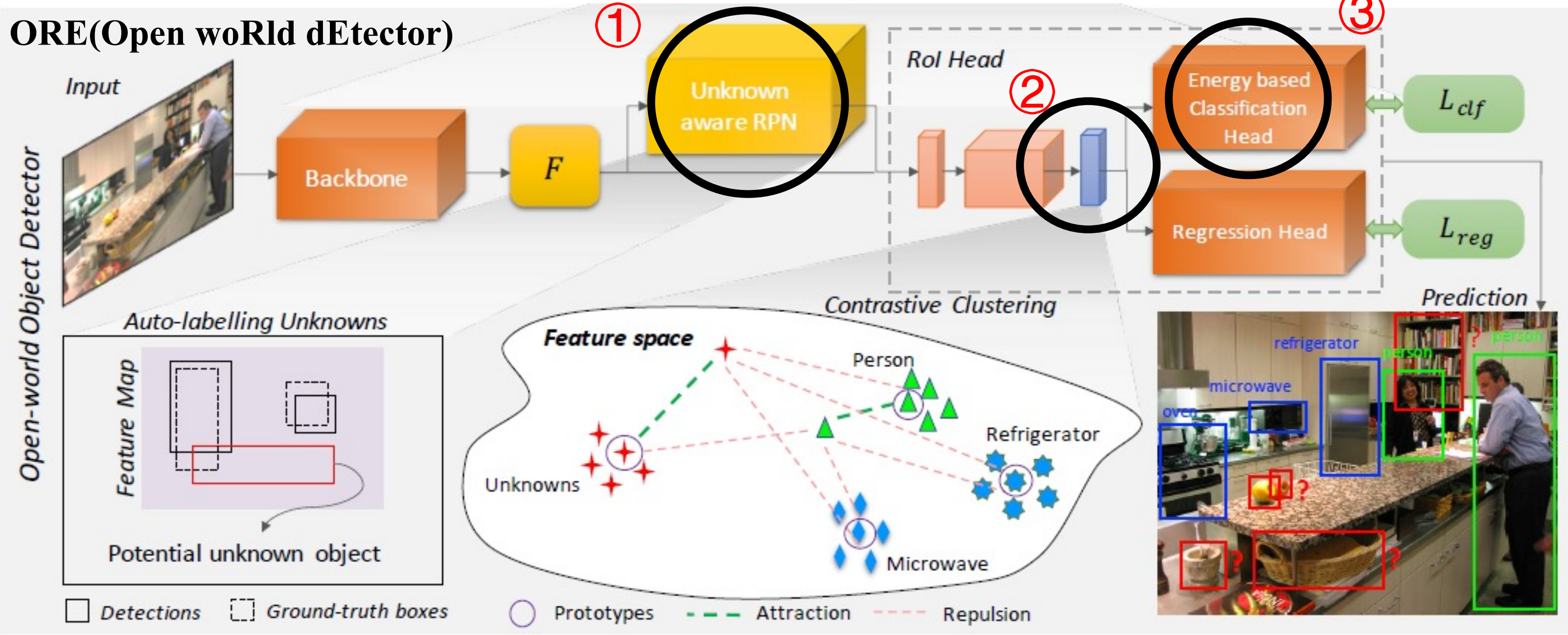
Compared with Classical Faster-RCNN:



(*) The article has been cited for **61** times until 05/14/2022.

2. Recap of *Towards Open World Object Detection*

Towards Open World Object Detection**,** KJ Joseph et al, IIT, CVPR2021 Oral **(Original OWOD)**



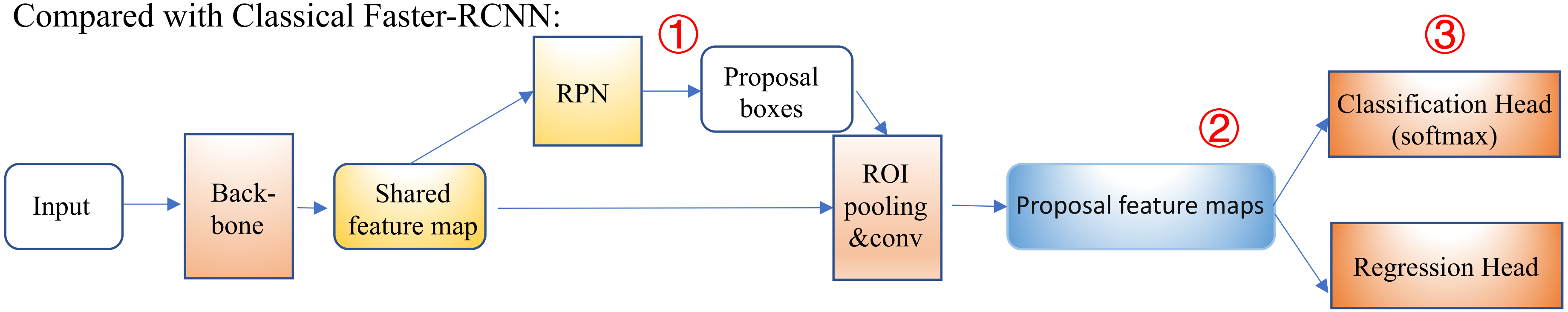
Note :

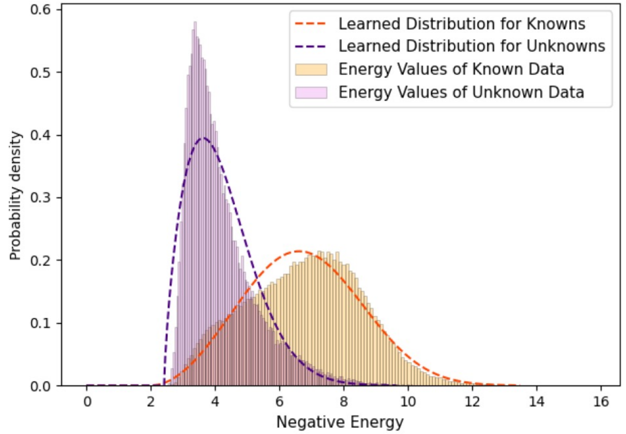
(1) OWOD directly uses an off-shelf incremental learning method after ORE.

(2) EBUI③ uses ***Weibull distribution*** to fit the known/unknown energy values, verified by a **held-out validation set**.

$\xi_{unk}(\boldsymbol{f})$, with a set of shifted Weibull distributions. These distributions were found to fit the energy data of a small held out validation set (with both knowns and unknowns instances) very well, when compared to Gamma, Exponential

Compared with Classical Faster-RCNN:





2. Recap of *Towards Open World Object Detection*

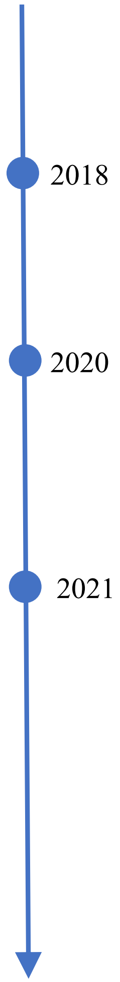
Question: What is the difference between **OWOD** and **Open-set Object Detection** ?

Paper2.1: ***Dropout Sampling for Robust Object Detection in Open-Set Conditions*** , D Miller et al, QUT, **ICRA 2018**
[Deal with label uncertainty via Dropout Sampling to reject previously unseen objects.**]** **(** **111** citations until 05/12/2022)

Paper2.2: ***The Overlooked Elephant of Object Detection: Open Set***, Akshay Dhamija et al, UCCS, **WACV 2020**
[**First** to formalize the issue of **open-set** object detection. **NO** new methods, conducting experiments on some SOTA **closed-set** detectors.**]** **(** **30** citations until 05/12/2022)

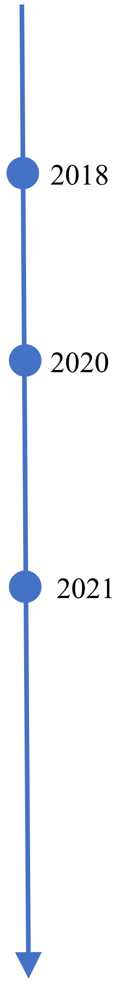
Paper2.3: ***Uncertainty for Identifying Open-Set Errors in Visual Object Detection***, D Miller et al, QUT, **RAL 2021**
(**7** citations until 05/12/2022)

……



2. Recap of *Towards Open World Object Detection*

Question: What is the difference between **OWOD** and **Open-set Object Detection** ?



Paper2.1: ***Dropout Sampling for Robust Object Detection in Open-Set Conditions*** , D Miller et al, QUT, **ICRA 2018**
[Deal with label uncertainty via Dropout Sampling to reject previously unseen objects.**]** **(** **111** citations until 05/12/2022)

Paper2.2: ***The Overlooked Elephant of Object Detection: Open Set***, Akshay Dhamija et al, UCCS, **WACV 2020**
[**First** to formalize the issue of **open-set** object detection. **NO** new methods, conducting experiments on some SOTA **closed-set** detectors.**]** **(** **30** citations until 05/12/2022)

Paper2.3: ***Uncertainty for Identifying Open-Set Errors in Visual Object Detection***, D Miller et al, QUT, **RAL 2021**
(**7** citations until 05/12/2022)

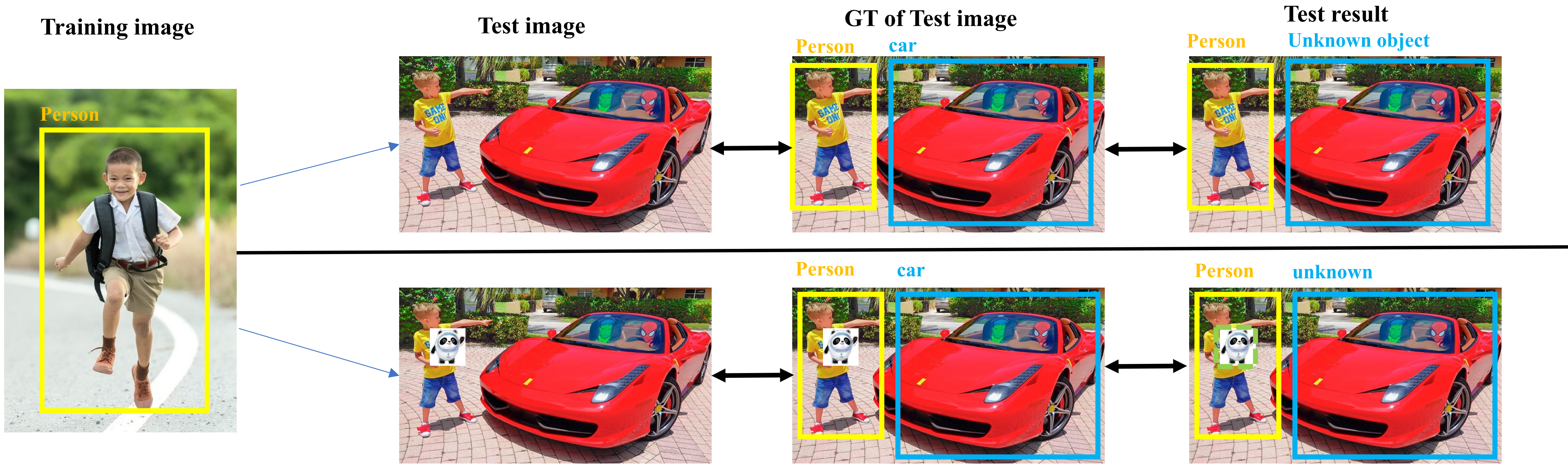
……

Conclusion: There is actually NO concrete gap between OWOD and Open-set Object Detection, both of them need an open-set detector.

2. Recap of *Towards Open World Object Detection*

Question: What is the difference between **OWOD** and **Open-set Object Detection** ?

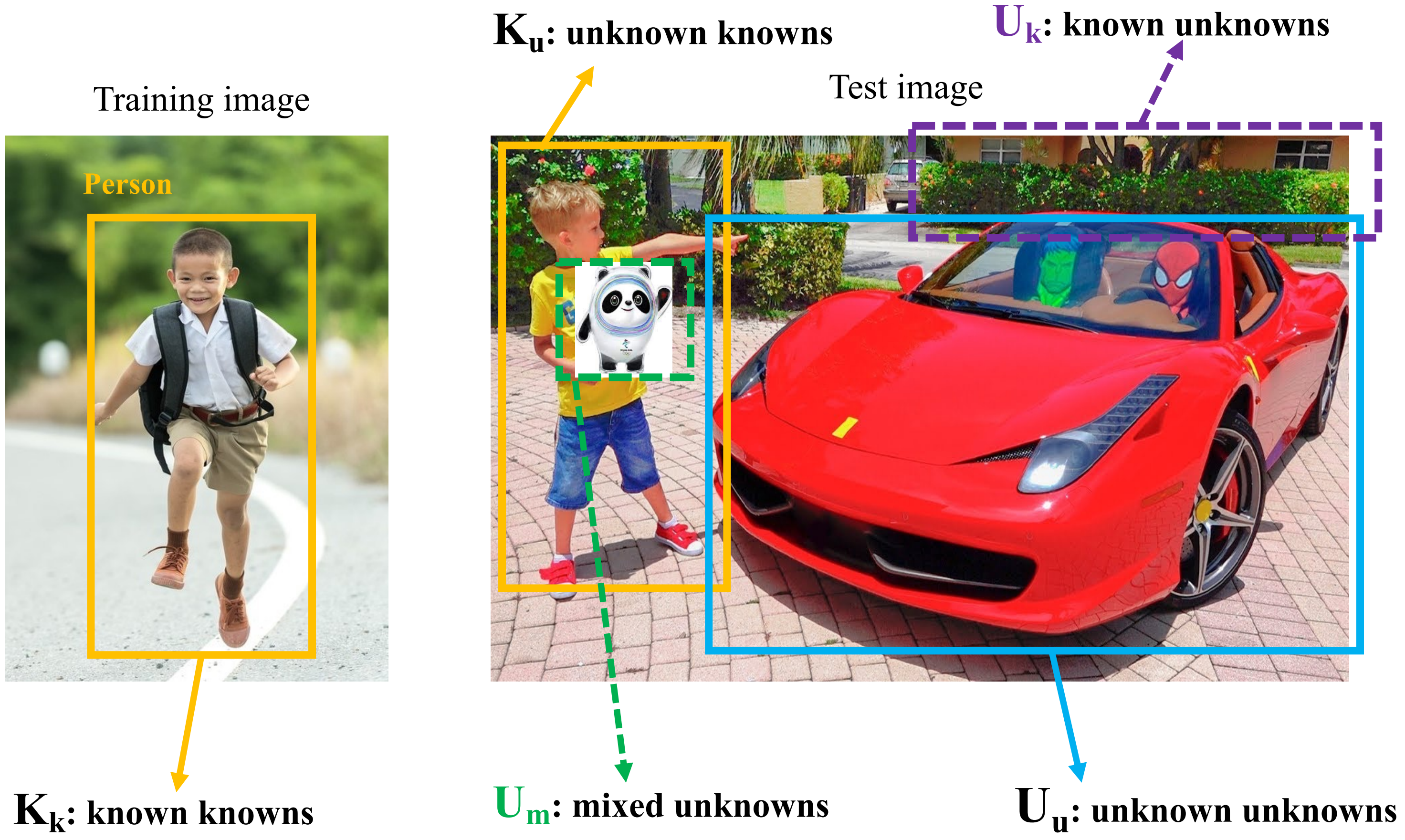
Paper2.2: ***The Overlooked Elephant of Object Detection: Open Set***, Akshay Dhamija et al, UCCS, WACV 2020



2. Recap of *Towards Open World Object Detection*

Question: What is the difference between **OWOD** and **Open-set Object Detection** ?

Paper2.2: ***The Overlooked Elephant of Object Detection: Open Set***, Akshay Dhamija et al, UCCS, WACV 2020



U_m refers to unlabelled objects in the dataset, but different to U_k which contains the authentic background or garbage objects, like sky, grass, trees.

Both U_m and U_k are **NOT** labelled in the dataset.

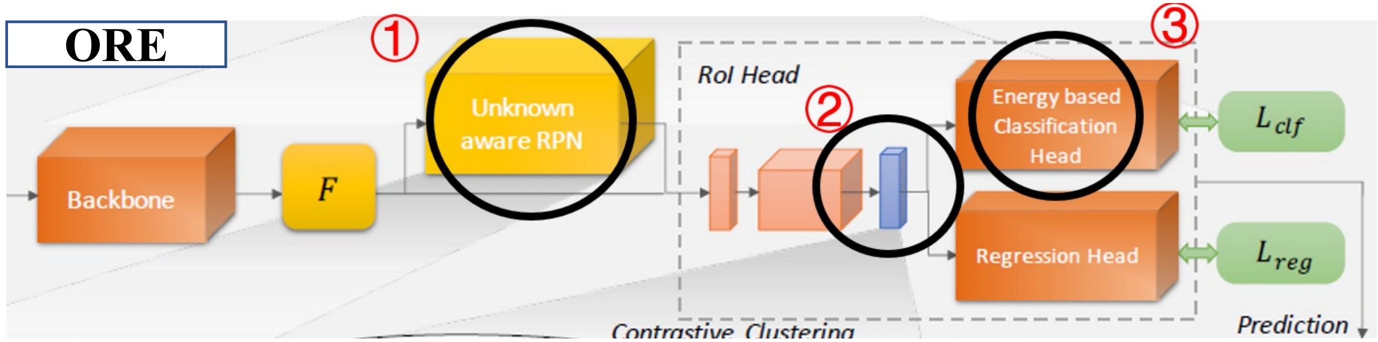
Article's view:

The presence of U_m **CANNOT** be 100% avoided, we can only reduce the number of U_m.

3.1 *Revisiting Open World Object Detection*

***Revisiting Open World Object Detection,** Xiaowei Zhao et al, Beihang Univeristy, preprint 2022.1

(1) Improvement in methods compared with ORE in OWOD (still based on Faster-RCNN)



A. Unknown aware RPN①

Drawback：
Trained on **labelled** known objects, unknown-aware-RPN may roll out some **authentic background regions** as unknown objects mistakenly, as shown below.

ORE

Ours

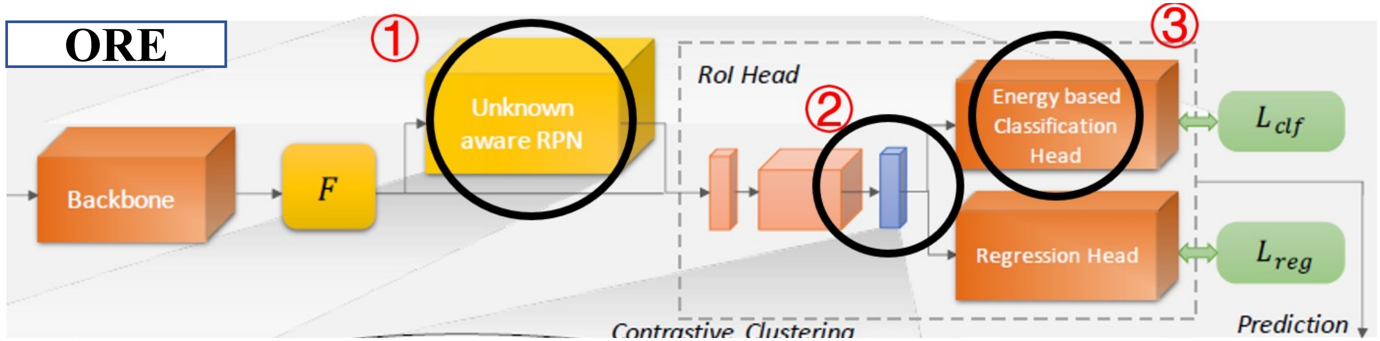
(*) The article was written in CVPR format, but not published by CVPR2022, submission unclear. Citation = 1

3.1 *Revisiting Open World Object Detection*

***Revisiting Open World Object Detection,** Xiaowei Zhao et al, Beihang Univeristy, preprint 2022.1

(1) Improvement in methods compared with ORE in OWOD (still based on Faster-RCNN)

A. Unknown aware RPN①



Drawback：
Trained on **labelled** known objects, unknown-aware-RPN may roll out some **authentic background regions** as unknown objects mistakenly, as shown below.

ORE

Ours

Solution:
Introduce an **Auxiliary Proposal Advisor**, which can be actually any unsupervised object detection approach, like **Selective Search.** The advisor helps to **confirm** the proposals produced by RPN.

$$\bar{\mathbf{S}}_i = \mathbf{S}_i \times \mathcal{I}\{\max_{1\le j\le|\widetilde{\mathbf{P}}^+|}(\mathrm{IOU}(\mathbf{P}_i^{(u)+}, \widetilde{\mathbf{P}}_j^+)) > \theta\},$$ (S: Objectness Score)

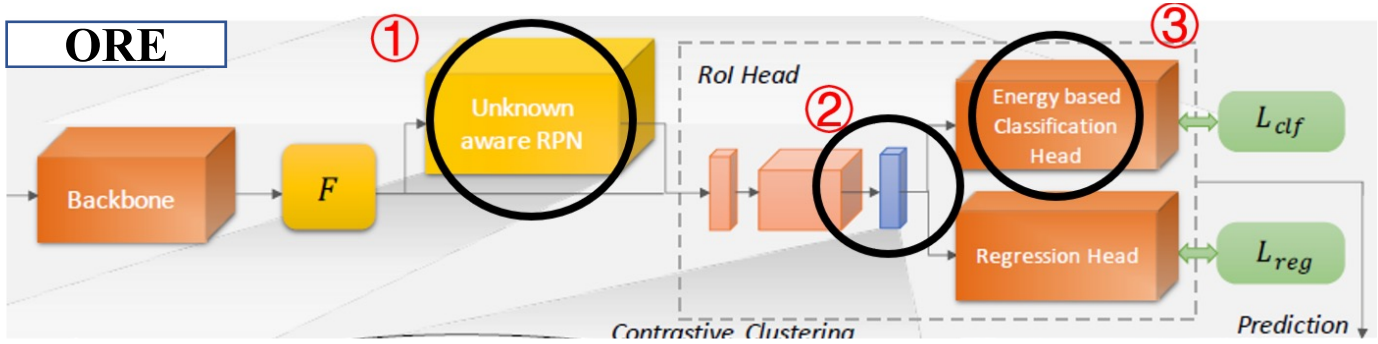
(*) The article was written in CVPR format, but not published by CVPR2022, submission unclear. Citation = 1

3.1 *Revisiting Open World Object Detection*

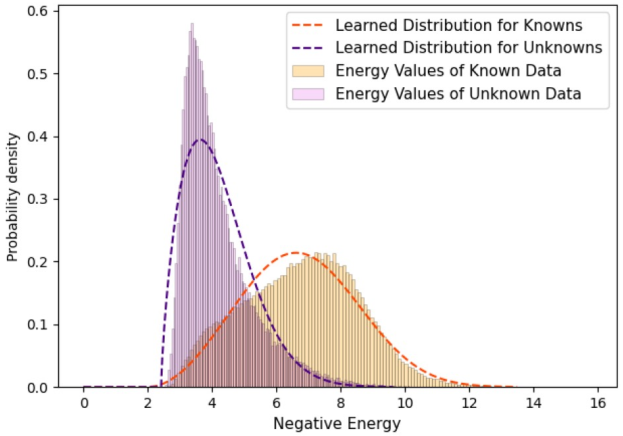
***Revisiting Open World Object Detection,** Xiaowei Zhao et al, Beihang Univeristy, preprint 2022.1

(1) Improvement in methods compared with ORE in OWOD (still based on Faster-RCNN)

B. Energy-based Unknown Identifier③



Drawback：
The UBUI is charged with its **data leakage** risk because the *Weibull Distribution* to fit the known/unknown energy values is chosen based on **a side validation set**(with annotations for unknown objects).

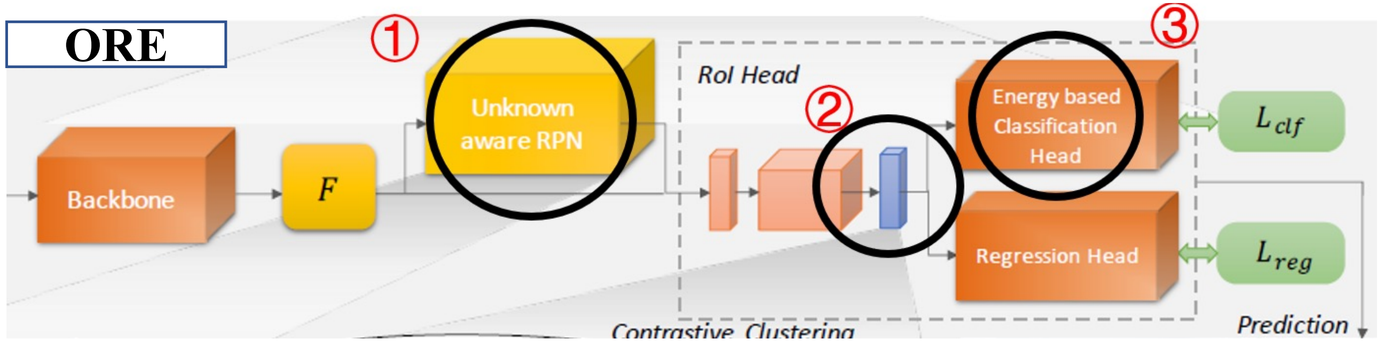


(*) The article was written in CVPR format, but not published by CVPR2022, submission unclear. Citation = 1

3.1 *Revisiting Open World Object Detection*

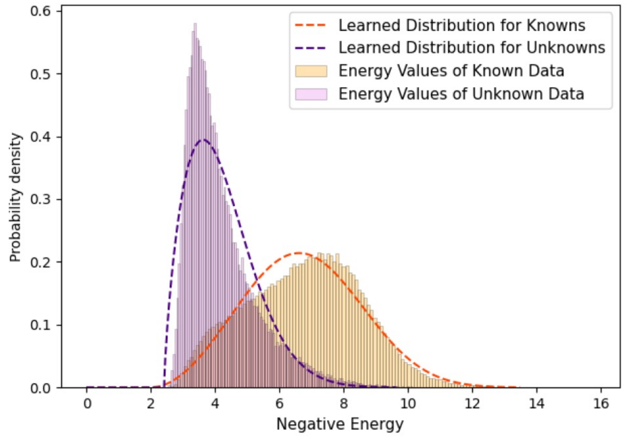
***Revisiting Open World Object Detection,** Xiaowei Zhao et al, Beihang Univeristy, preprint 2022.1

(1) Improvement in methods compared with ORE in OWOD (still based on Faster-RCNN)

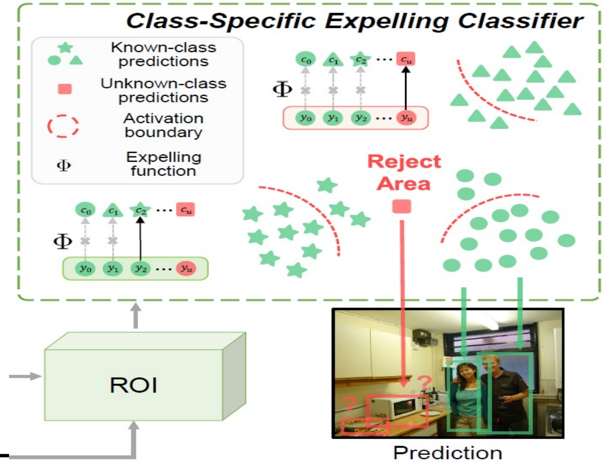


B. Energy-based Unknown Identifier③

Drawback：
The UBUI is charged with its **data leakage** risk because the *Weibull Distribution* to fit the known/unknown energy values is chosen based on **a side validation set**(with annotations for unknown objects).



Solution: Class-Specific Expelling Classifier



(*) The article was written in CVPR format, but not published by CVPR2022, submission unclear. Citation = 1

3.1 *Revisiting Open World Object Detection*

Revisiting Open World Object Detection, Xiaowei Zhao et al, Beihang Univeristy, preprint 2022.1

(2) New Benchmark Protocols for Open-World Object Detection

√ **: principle already met by OWOD**
× **: principle NOT met yet by OWOD**

The new benchmark meets 5 **principles** list below:

A. √ **Class Openness :** During testing, instances from known classes and unknown classes all appear.

B. √ **Task Increment :** Known classes are increasing in size, the task is thus incrementally developed.

C. × **Annotation Specificity :**
For **training** and **validation** set, only labels of **known** classes are assigned.
For **test** set, labels of both known and unknown classes are assigned, all novel classes as unseen.

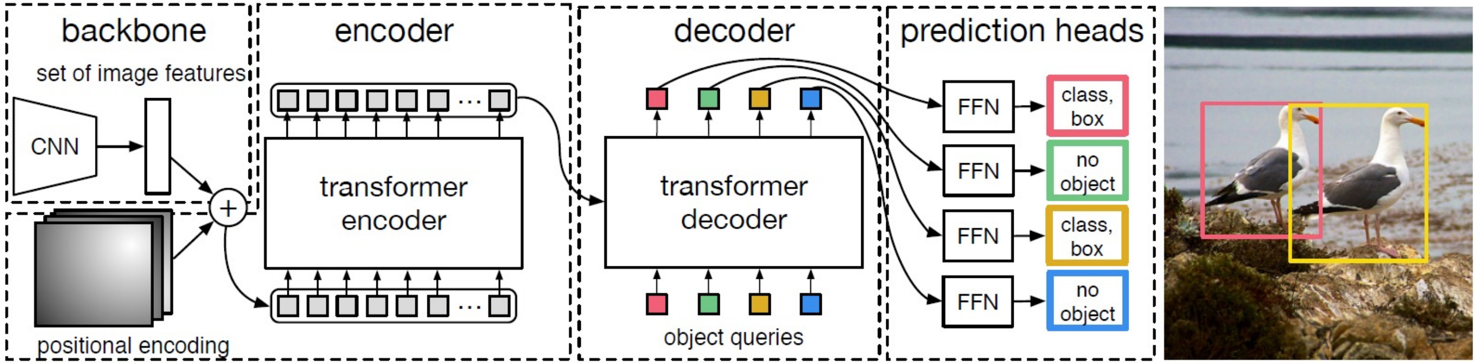
D. × **Label Integrity**
Ask for a **fully-annotated dataset** for testing, avoiding **false-positive mistake**s.

E. × **Data Specificity**
Ask for **NO** duplication inside dataset for testing.
e.g., sofa as known class, and furniture as unknown class.

(3) New metrics for Evaluation of Open-World Object Detection models. (skip over here)

3.2 Transformer-based OWOD(**OW-DeTR**)

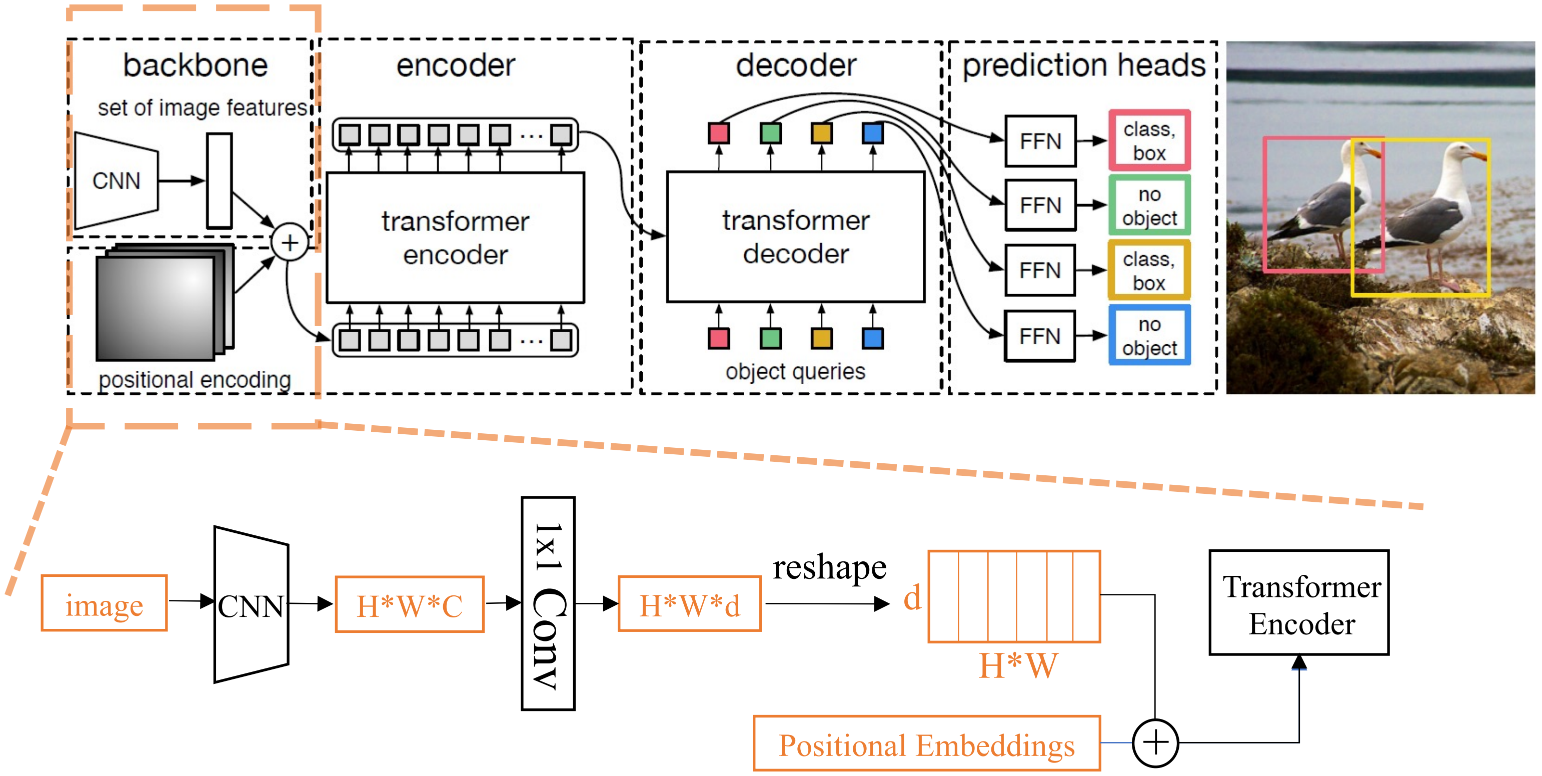
3.2.1 * DeTR: A Transformer-based Object Detector



(*) *End-to-End Object Detection with Transformers*, Nicolas Carion et al, FAIR, ECCV 2020

3.2 Transformer-based OWOD(**OW-DeTR**)

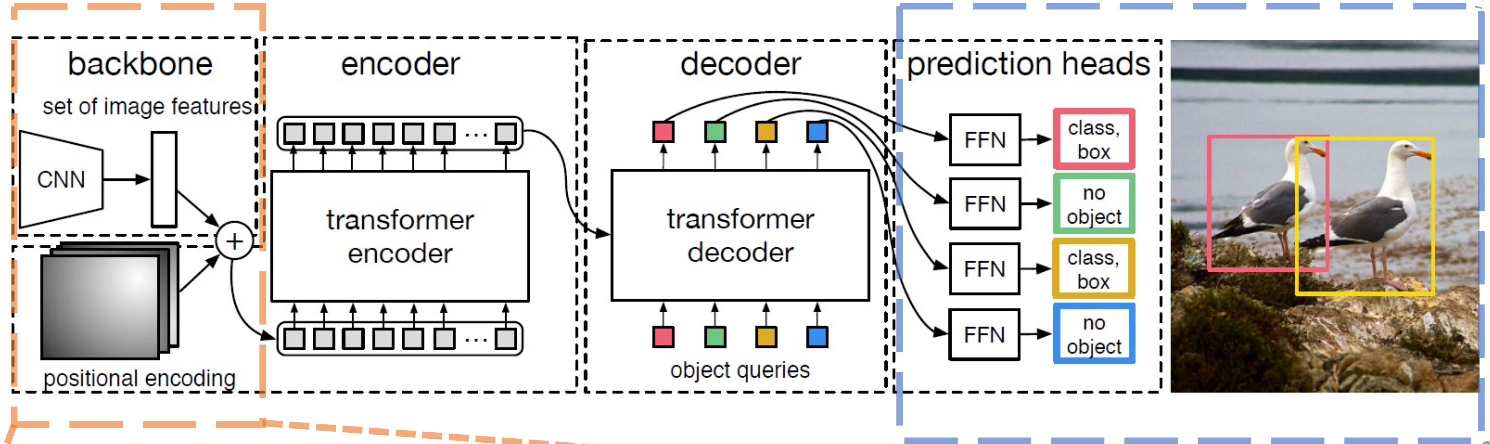
3.2.1 * **DeTR**: A Transformer-based Object Detector

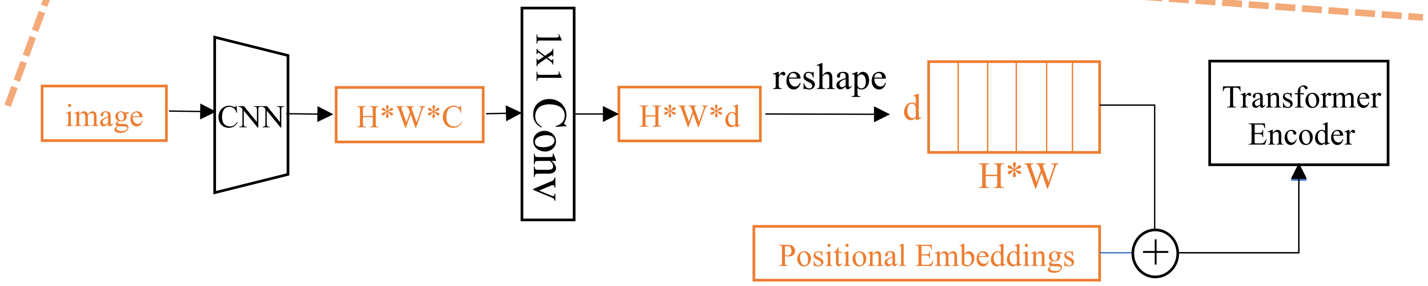


(*) *End-to-End Object Detection with Transformers*, Nicolas Carion et al, FAIR, ECCV 2020

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.1 * **DeTR**: A Transformer-based Object Detector

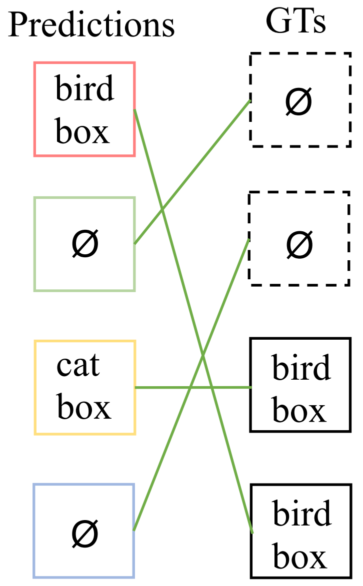




(1) FFN (Prediction Feed-Forward Network) :

FFN : MLP and a linear projection layer to produce predictions for classes and bounding boxes.
*Fixed output numbers: **N**

(2) Pair-wise Matching :



Bipartite Matching

σ : **pair-wise matching policy**

$$\hat{\sigma} = \arg\min_{\sigma \in \mathfrak{S}_N} \sum_i^N \mathcal{L}_{\mathrm{match}}(y_i, \hat{y}_{\sigma(i)})$$

(y : GT, y^hat : prediction)

$\mathcal{L}_{\mathrm{match}}$: pair-wise matching cost
$= \mathbf{L}_{\mathbf{class}} + \mathbf{L}_{\mathbf{box}}$

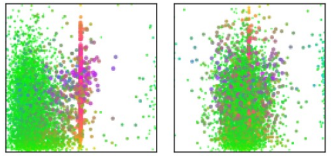
NOTE: The matching policy σ is efficiently calculated by *Hungarian Algorithm*.

(*) *End-to-End Object Detection with Transformers*, Nicolas Carion et al, FAIR, ECCV 2020

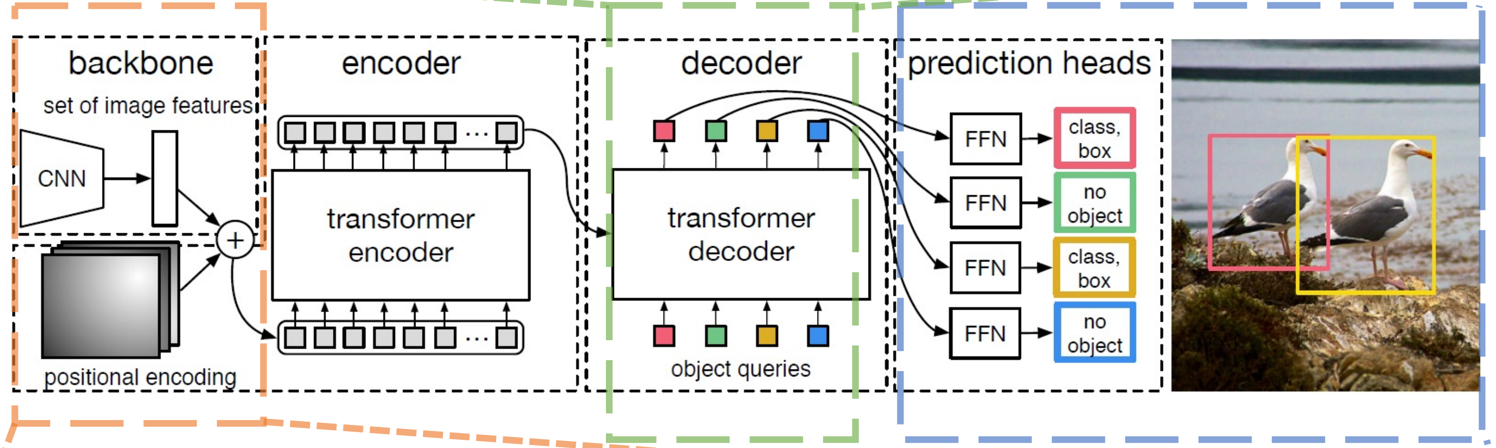
3.2 Transformer-based OWOD(**OW-DeTR**)

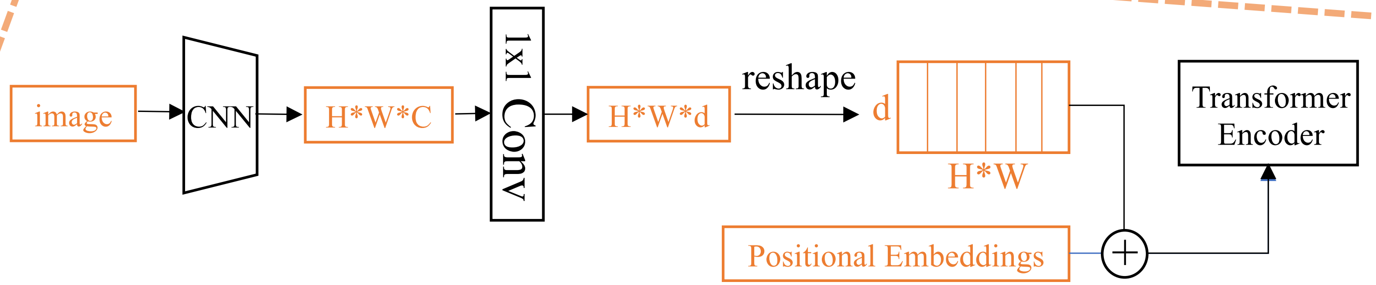
3.2.1 * **DeTR**: A Transformer-based Object Detector

Object Queries: Learnable Positional Embeddings, fixed numbers as **N**.
(1) **N** refers to the maximum number of object predictions of the model.
(2) Each query 'focus' on some specific region of the image.
(3) Randomly initialized.



: small bboxes
: vertical bboxes
: horizontal bboxes

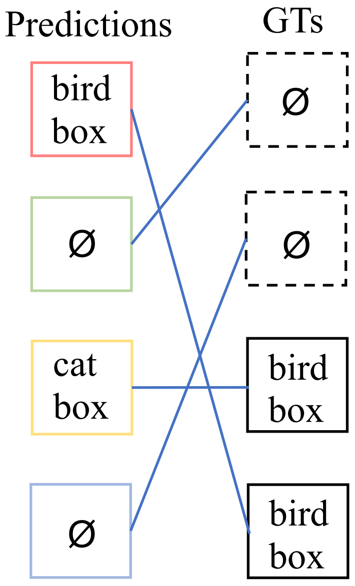




(1) FFN (Prediction Feed-Forward Network) :

FFN : MLP and a linear projection layer to produce predictions for classes and bounding boxes.
*Fixed output numbers: **N**

(2) Pair-wise Matching :



Bipartite Matching

σ **: pair-wise matching policy**

$$\hat{\sigma} = \arg\min_{\sigma \in \mathfrak{S}_N} \sum_{i}^{N} \mathcal{L}_{\mathrm{match}}(y_i, \hat{y}_{\sigma(i)})$$

(y : GT, y^hat : prediction)

$\mathcal{L}_{\mathrm{match}}$: pair-wise matching cost
$= \mathbf{L}_{\mathbf{class}} + \mathbf{L}_{\mathbf{box}}$

NOTE: The matching policy σ is efficiently calculated by *Hungarian Algorithm*.

(*) *End-to-End Object Detection with Transformers*, Nicolas Carion et al, FAIR, ECCV 2020

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ***OW-DeTR: Open-world Detection Transformer**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

Improvements of **OW-DeTR** compared with original **OWOD**:

(1) An **end-to-end** framework for Open-World Object Detection. (Incremental Learning still left outside)

(2) Abandon the held-out validation for unknown identifier, **avoiding data leakage**.

(3) Considering **background VS foreground** to better constitute a **valid object**.

(4) Attention-driven **pseudo labeling** overcomes the bias caused by unknown-aware RPN

(5) Better **performance** on Open-World Object Detection tasks.

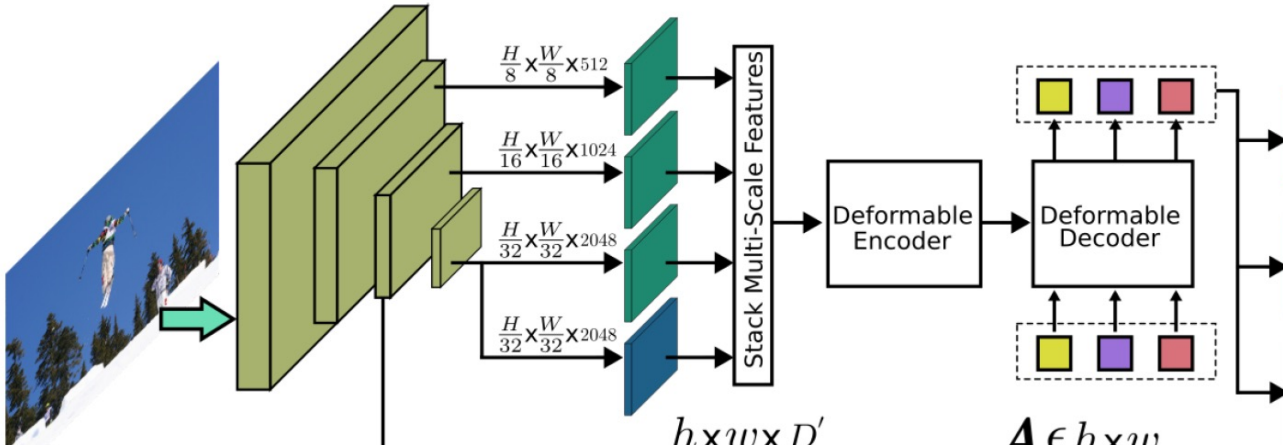
Note:
The **OW-DeTR** is based on the **Deformable DeTR(DDeTR)** instead of the **original DeTR**, yet the deformable one is just with minor changes to the original one.

(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ***OW-DeTR: Open-world Detection Transformer**,* Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(1) Architecture of OW-DeTR

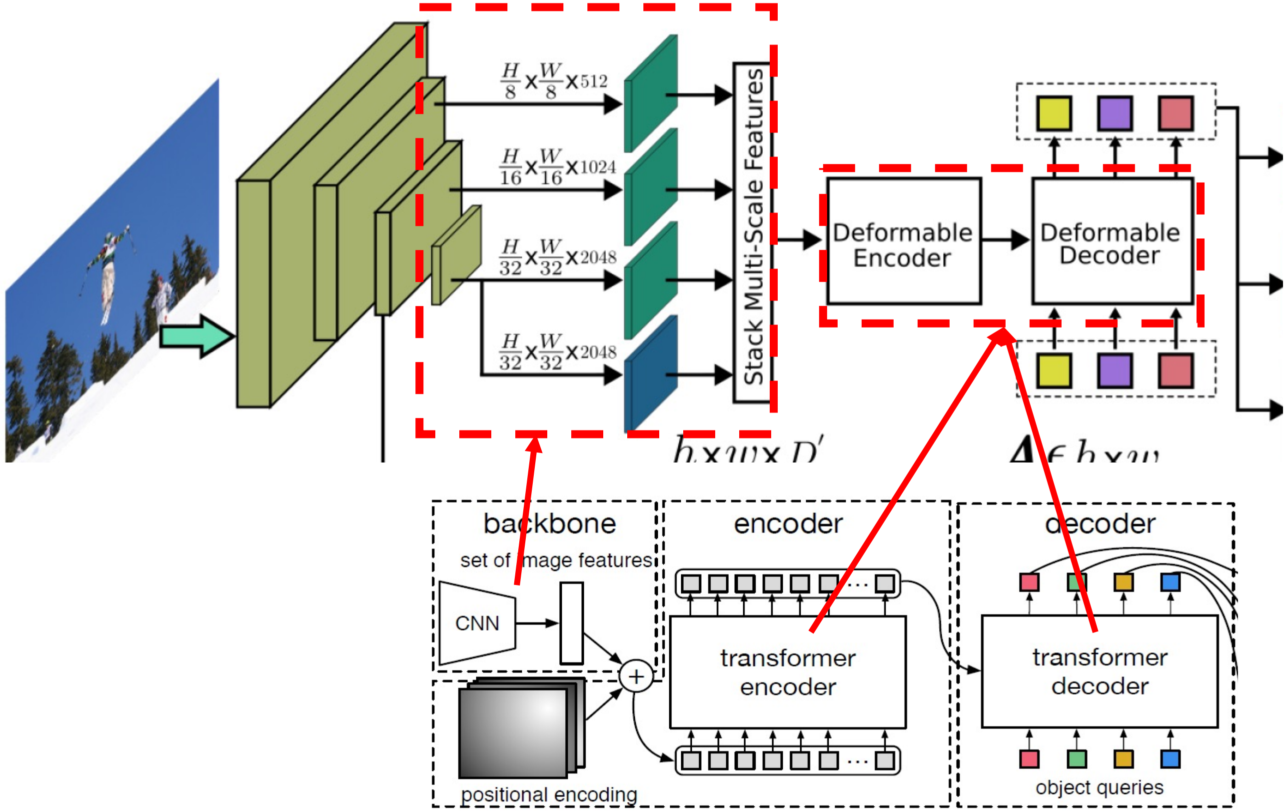


(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 **_*OW-DeTR: Open-world Detection Transformer_**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(1) Architecture of OW-DeTR



What is new in DDeTR vs DeTR

(1) Multi-scale Context

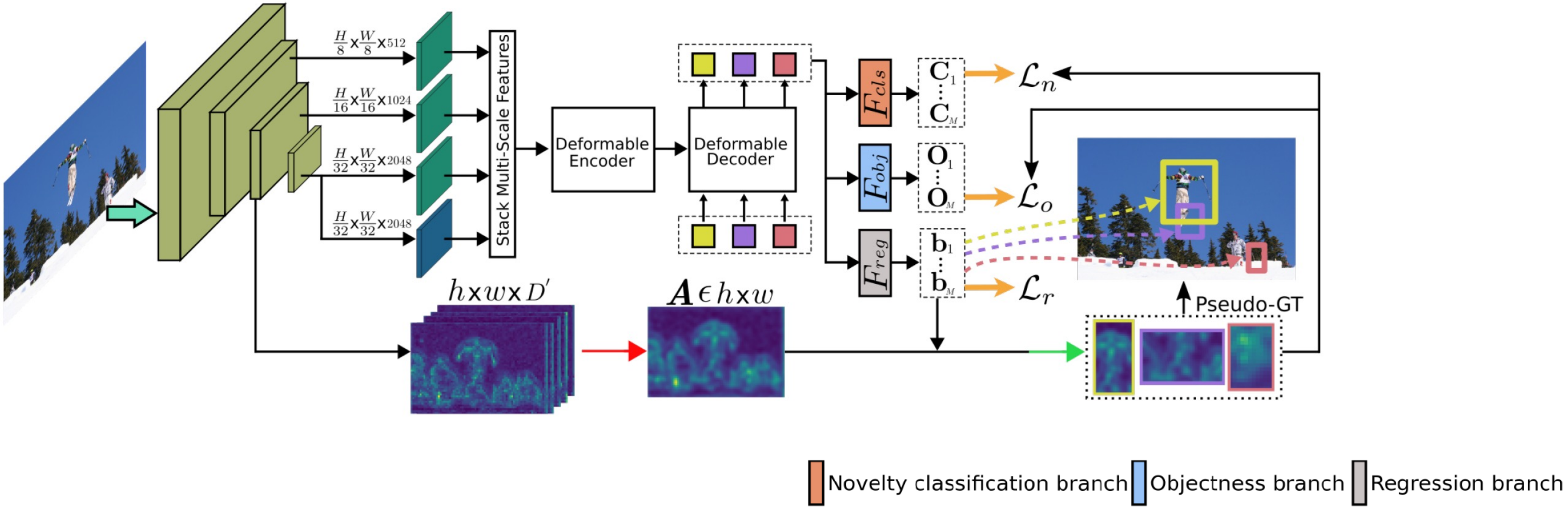
(2) Deformable encoder/decoder

(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ******OW-DeTR: Open-world Detection Transformer*, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(1) Architecture of OW-DeTR

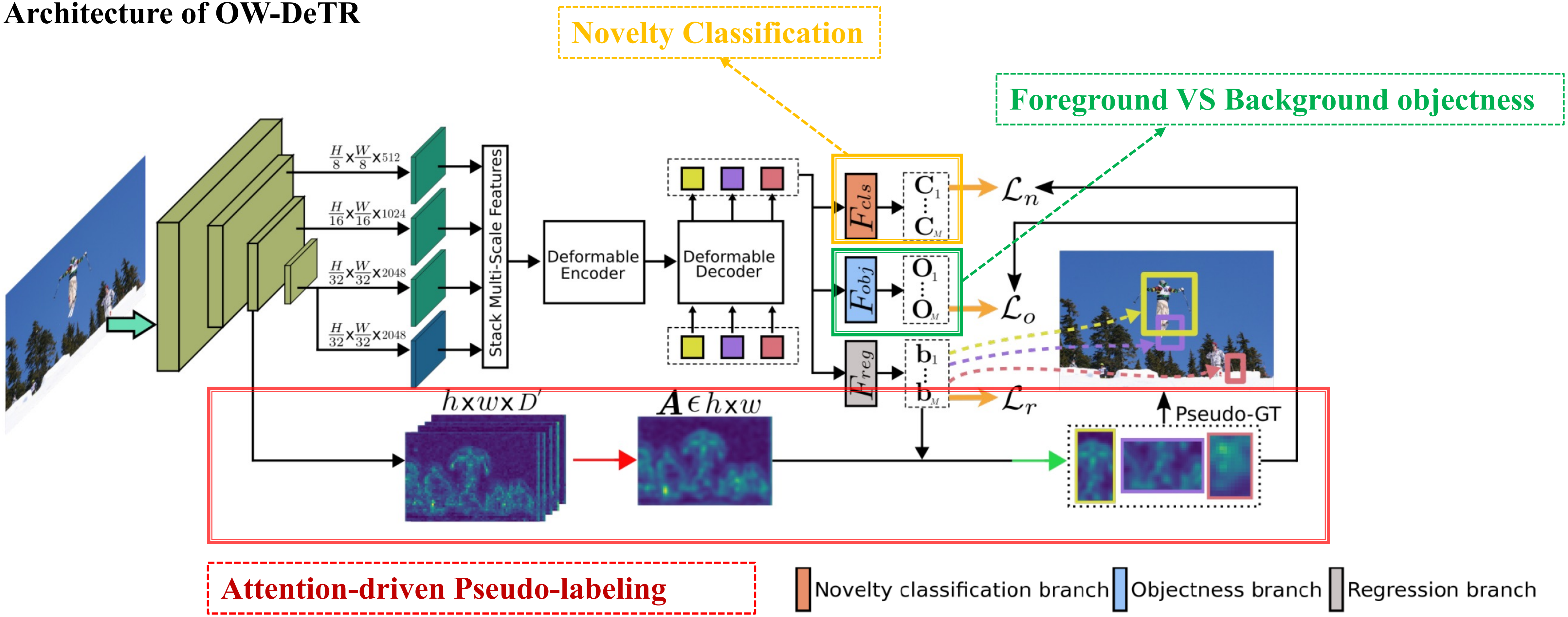


(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 **_*OW-DeTR: Open-world Detection Transformer_**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(1) Architecture of OW-DeTR

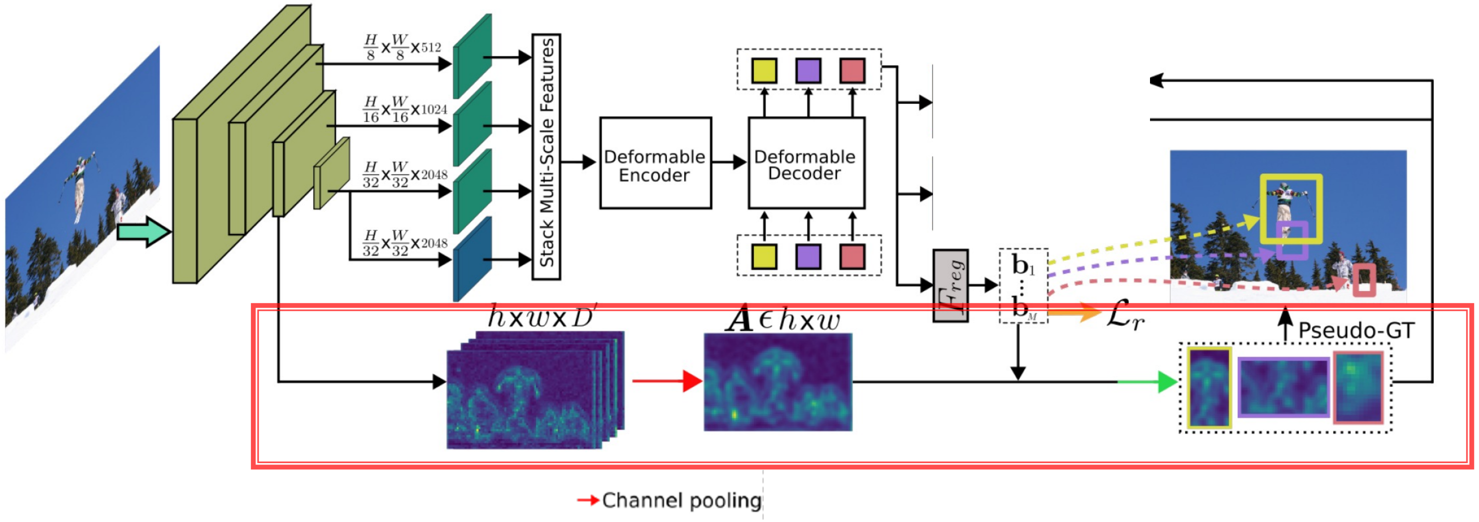


(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ***_OW-DeTR: Open-world Detection Transformer_**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(2) Attention-driven Pseudo Labeling

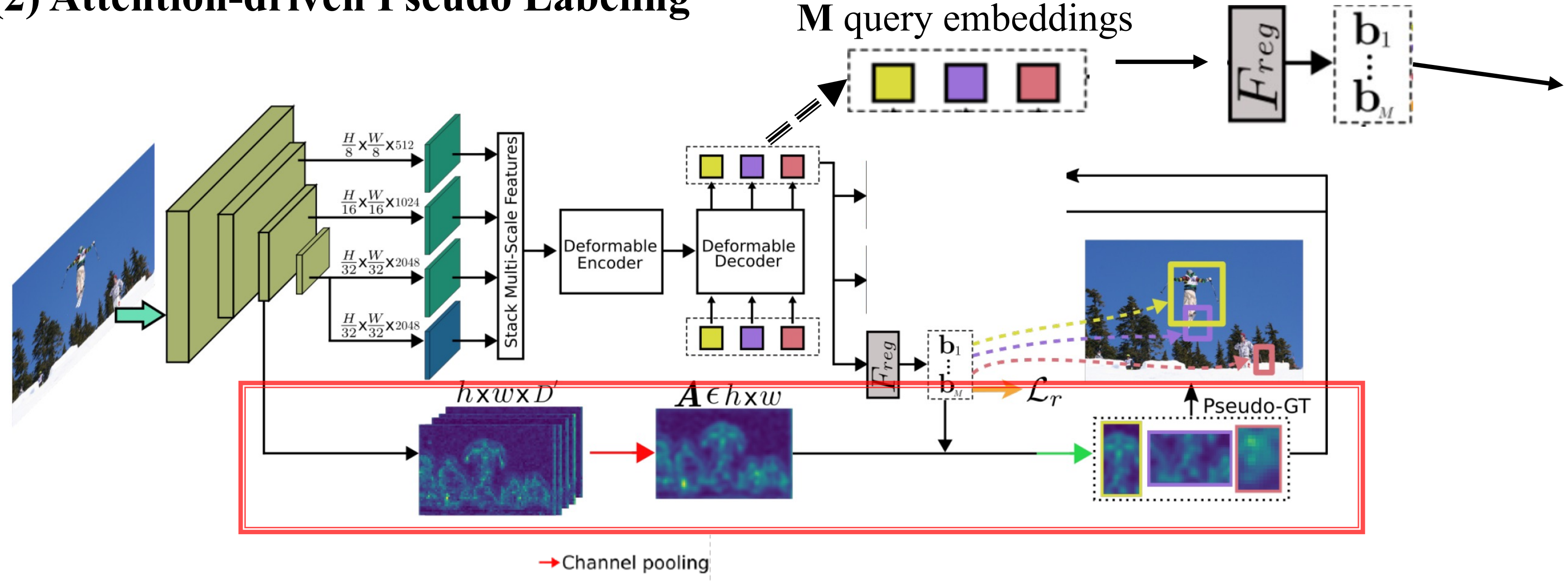


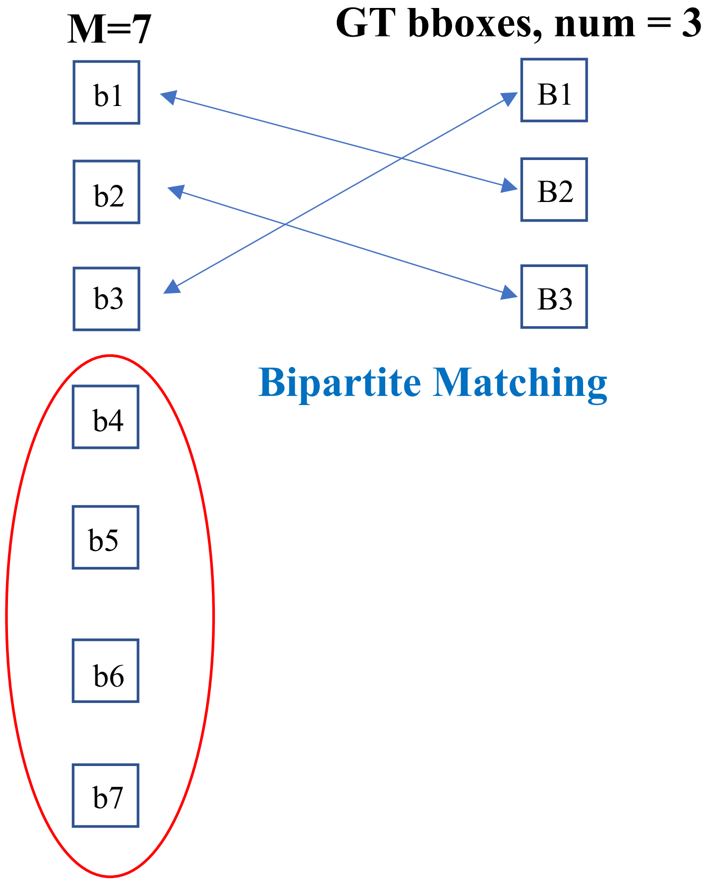
(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ******OW-DeTR: Open-world Detection Transformer*, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(2) Attention-driven Pseudo Labeling



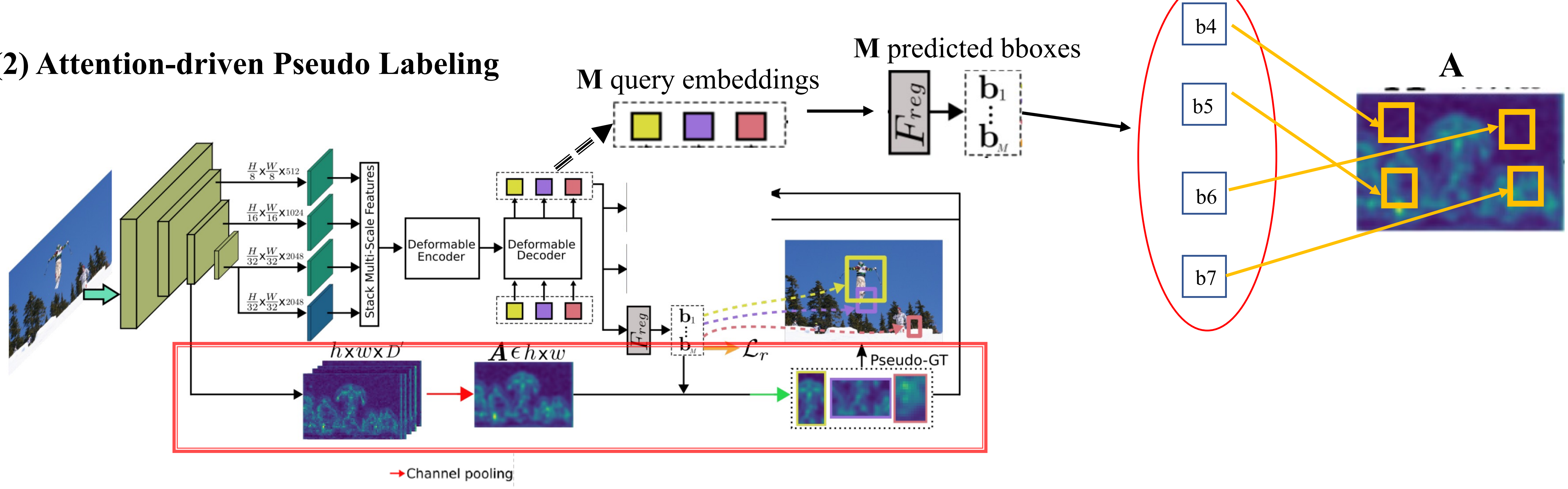


(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 **_*OW-DeTR: Open-world Detection Transformer_**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(2) Attention-driven Pseudo Labeling

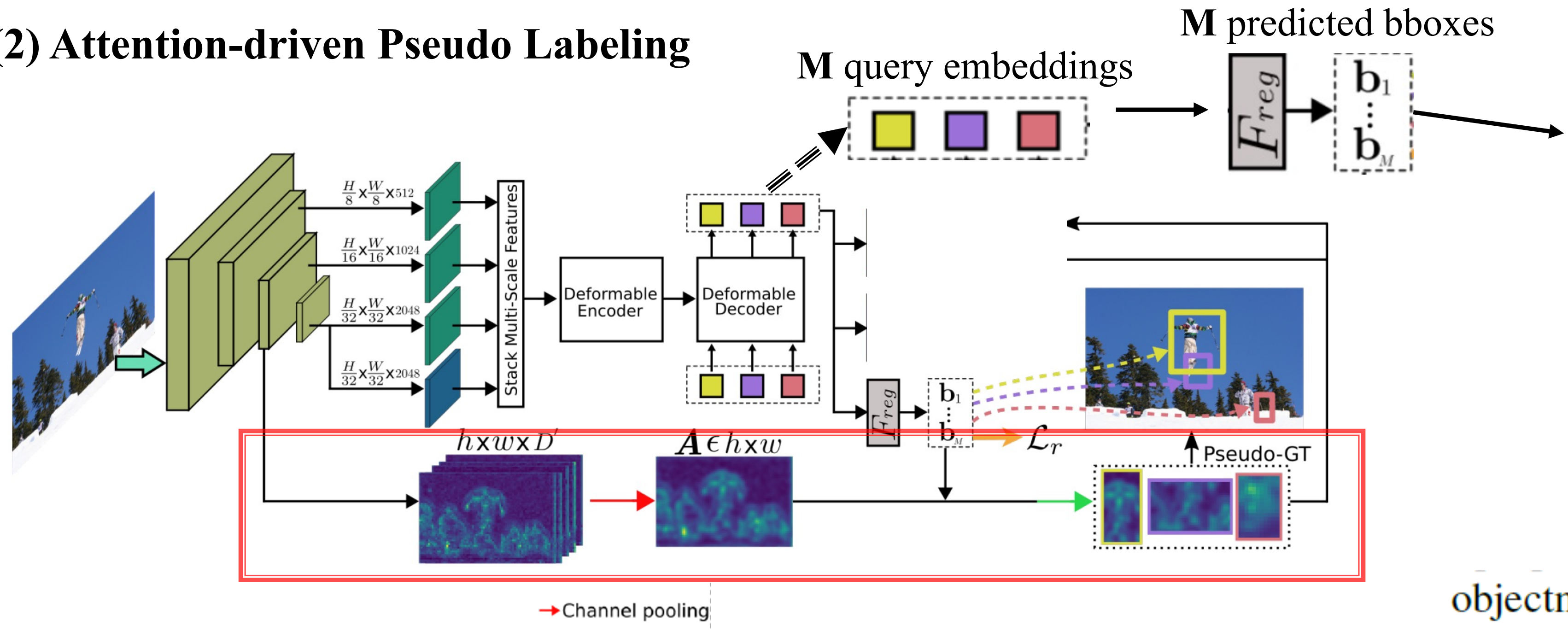


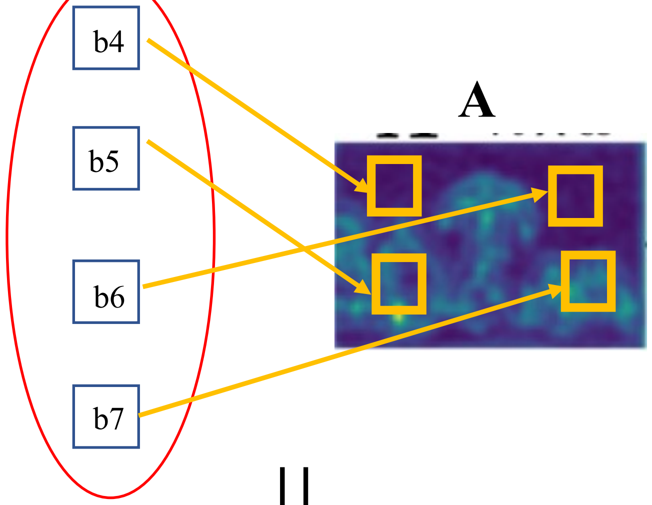
(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 **OW-DeTR: Open-world Detection Transformer,* Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(2) Attention-driven Pseudo Labeling





$$\text{objectness score } s_o(\boldsymbol{b}) = \frac{1}{h_b \cdot w_b} \sum_{x_b-\frac{w_b}{2}}^{x_b+\frac{w_b}{2}} \sum_{y_b-\frac{h_b}{2}}^{y_b+\frac{h_b}{2}} \boldsymbol{A},$$

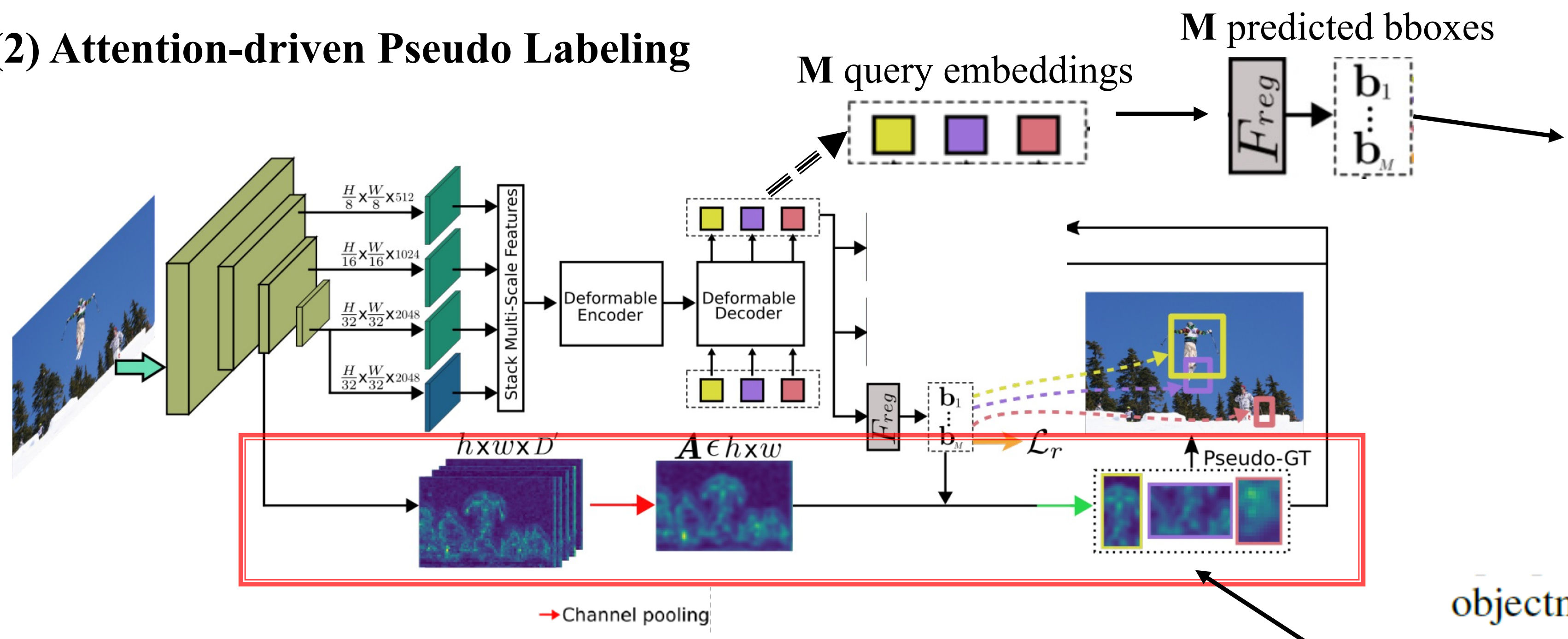
s(b4) = 0.32, s(b5) = 0.31
s(b6) = 0.28, s(b7) = 0.53

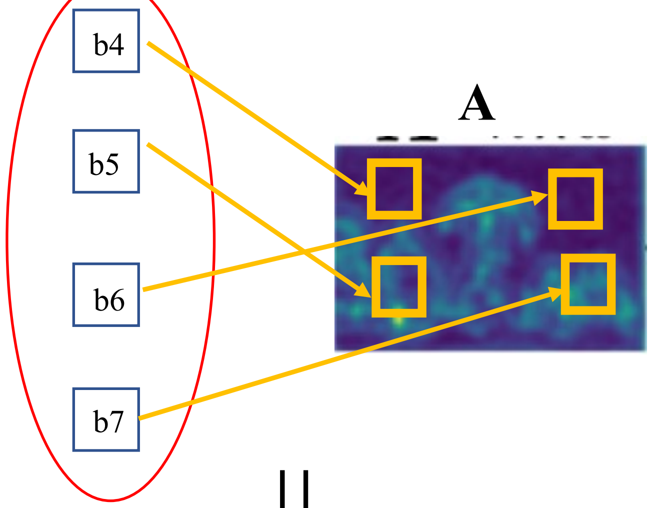
(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ***OW-DeTR: Open-world Detection Transformer*,** Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(2) Attention-driven Pseudo Labeling





$$\text{objectness score } s_o(\boldsymbol{b}) = \frac{1}{h_b \cdot w_b} \sum_{x_b - \frac{w_b}{2}}^{x_b + \frac{w_b}{2}} \sum_{y_b - \frac{h_b}{2}}^{y_b + \frac{h_b}{2}} \boldsymbol{A},$$

s(b4) = 0.32, s(b5) = 0.31
s(b6) = 0.28, s(b7) = 0.53

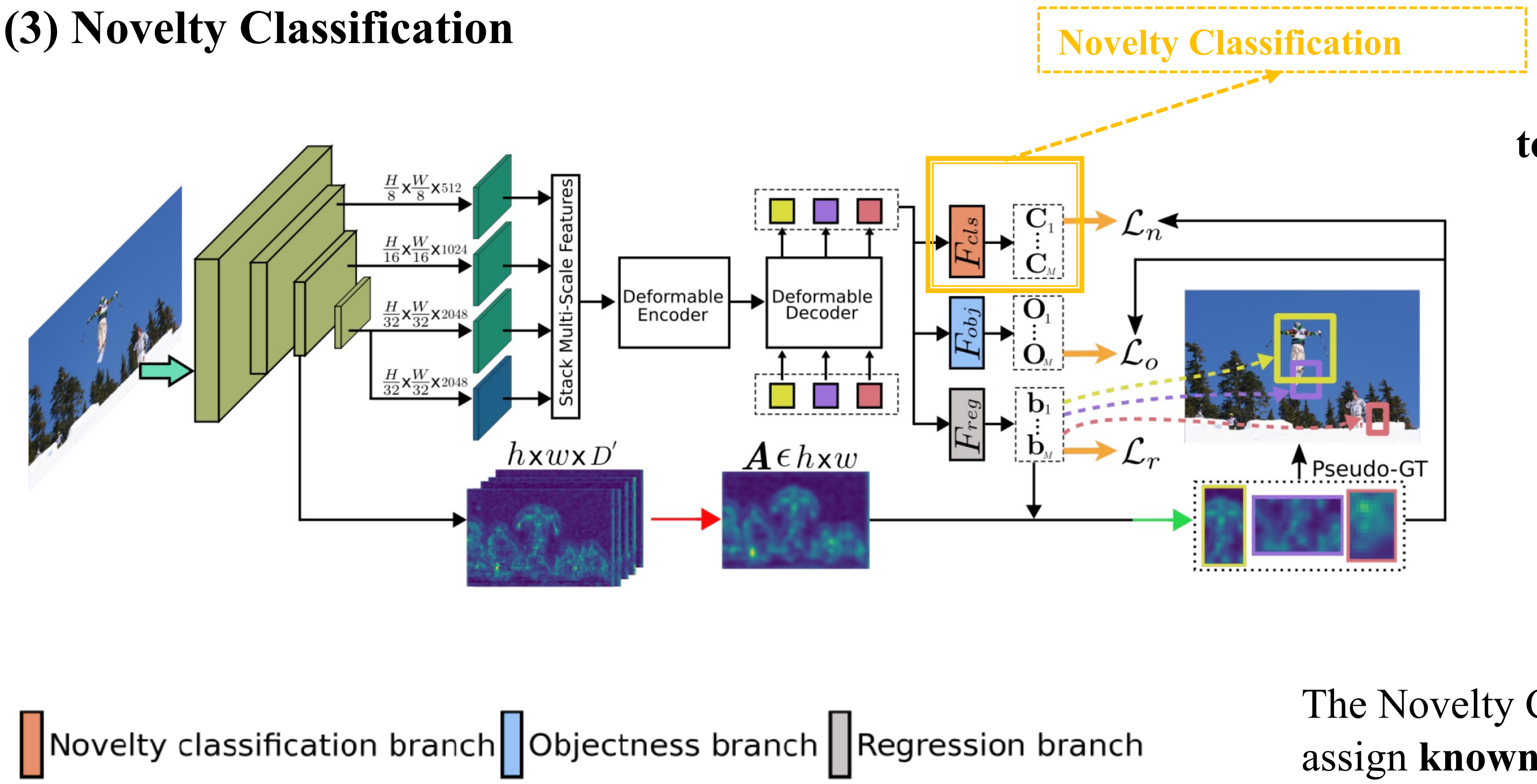
Rank them,
then select **top-K** as pseudo unknown object regions

(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 **_*OW-DeTR: Open-world Detection Transformer_**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(3) Novelty Classification



top-K selected pseudo unknown object regions

↓

K corresponding query embeddings $\{q_e^u\}$

↓

Assign label *unknown* to these $\{q_e^u\}$

↓

The Novelty Classification branch is trained to assign **known object region embeddings** to the known labels, and $\{\mathbf{q_e^u}\}$ to ***unknown*** .

LOSS: L_n

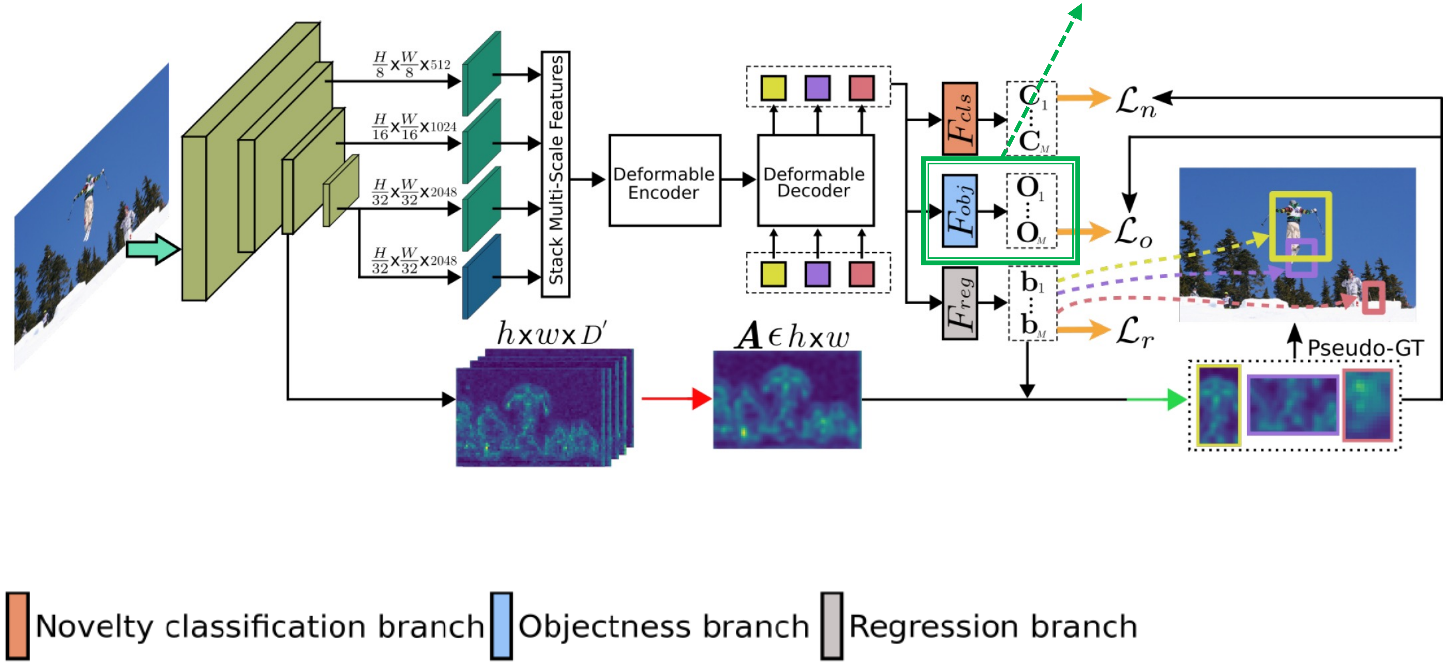
(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 ***OW-DeTR: Open-world Detection Transformer,*** Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(4) Objectness Branch

Foreground VS Background objectness



Score the query embeddings:

$$F_{obj} : \mathbb{R}^D \rightarrow [0, 1]$$

Known and unknowns objects → foreground

Background(no objects) → background

foreground: objectness score → **1**

LOSS: L_o

*** This part wasn't clearly explained in the article, citing the *Focal Loss*, proposed by**

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017. 5, 8

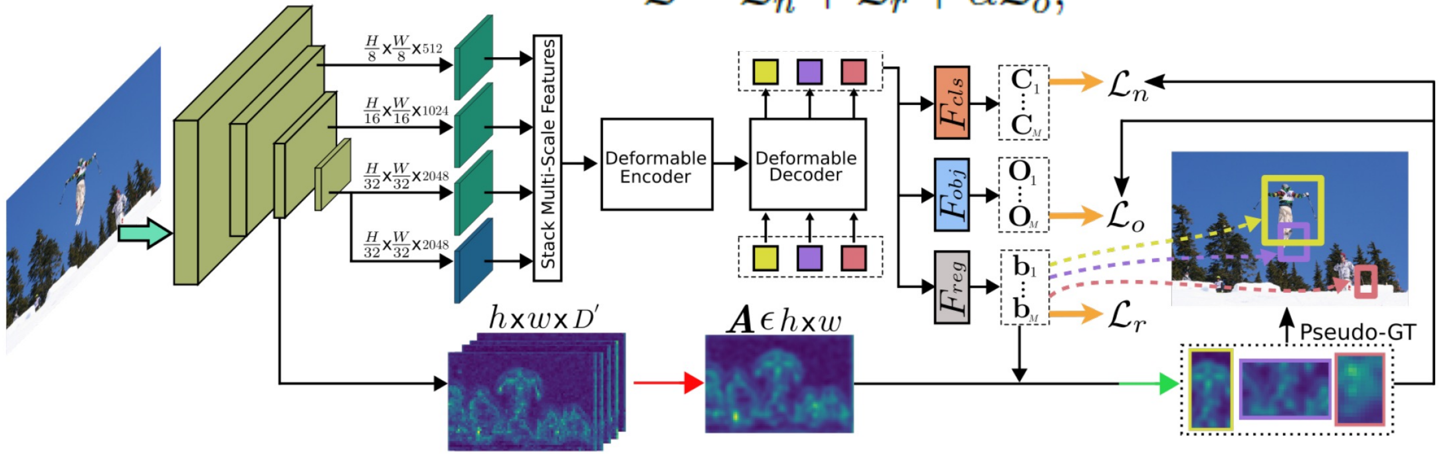
(*) This paper shares some same authors as original OWOD.

3.2 Transformer-based OWOD(**OW-DeTR**)

3.2.2 **OW-DeTR: Open-world Detection Transformer**, Akshita Gupta and K J Joseph et al, IIT, CVPR 2022

(5) Conclusion of OW-DeTR

$$\mathcal{L} = \mathcal{L}_n + \mathcal{L}_r + \alpha\mathcal{L}_o,$$



Novelty classification branch | Objectness branch | Regression branch

Methods:

Some new modules added to DDeTR;

highlight: Attention-driven Pseudo Labeling

Experiments & Evaluation:

similar to original OWOD

Same **datasets (**Pascal VOC, MSCOCO**)**

Same **metrics (**WI,A-OSE,mAP**)**

(*) This paper shares some same authors as original OWOD.

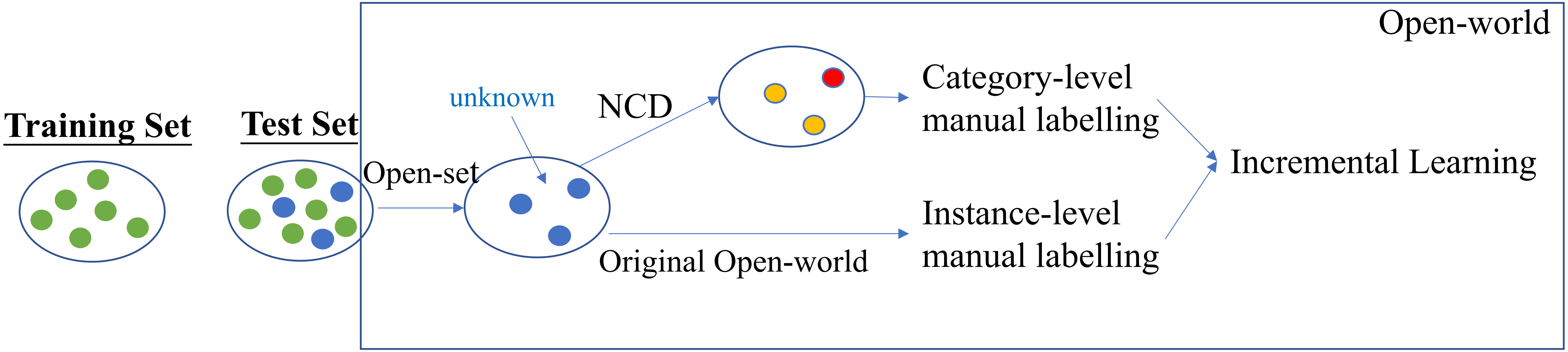
4. Novel Class Discovery(NCD)

4.1 Setting

Take Classification for example:

: instances belong to **known** labels

: instances belong to **unknown** labels



4. Novel Class Discovery(NCD)

4.2 NCD for classification

Paper4.1: *Learning to Discover Novel Visual Categories via Deep Transfer Clustering*(**DTC**)，Kai Han et al, ICCV 2019

Paper4.2: *Automatically Discovering and Learning Novel Visual Categories*(**AutoNovel**), Kai Han et al, ICLR 2020
[Improvement of **DTC**]

Paper4.3: *Novel Visual Category Discovery with Dual Ranking Statistics and Mutual Knowledge Distillation*,
Bingchen Zhao and Kai Han et al, NIPS 2021
[Improvement of **AutoNovel**]

Paper4.4: *Generalized Category Discovery*, Sagar Vaze and Kai Han et al, CVPR 2022

Paper4.5: *Spacing Loss for Discoverying Novel Categories*, KJ Joseph and Kai Han et al, CVPRW 2022

Note: KJ Joseph is the first auther of original OWOD,CVPR2021

4. Novel Class Discovery(NCD)

4.3 Open-world Classification + NCD

Paper4.6: *Open-World Semi-supervised Learning*(**OWSSL**), Kaidi Cao et,al, Stanford, ICLR 2022

4.4 Open-set Object Detection + NCD

Paper4.7: *Towards Open-set Object Detection and Discovery*, Jiayng Zheng et,al, ANU, CVPRW 2022

Open-set Object Detection
similar to ORE in OWOD

Novel Class Discovery
similar to DTC, a classical method of NCD