Workshop Safety & AI – DataIA
23/09/2020

VULNERABILITY OF PERSON RE-IDENTIFICATION MODELS TO METRIC ADVERSARIAL ATTACKS

CEA-List,
Université Paris-Saclay,
Vision and Learning Lab for Scene Analysis

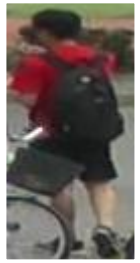
Quentin Bouniot, Romaric Audigier, Angélique Loesch

PERSON RE-IDENTIFICATION

Quentin Bouniot, Romaric Audigier, Angélique Loesch

PERSON RE-IDENTIFICATION

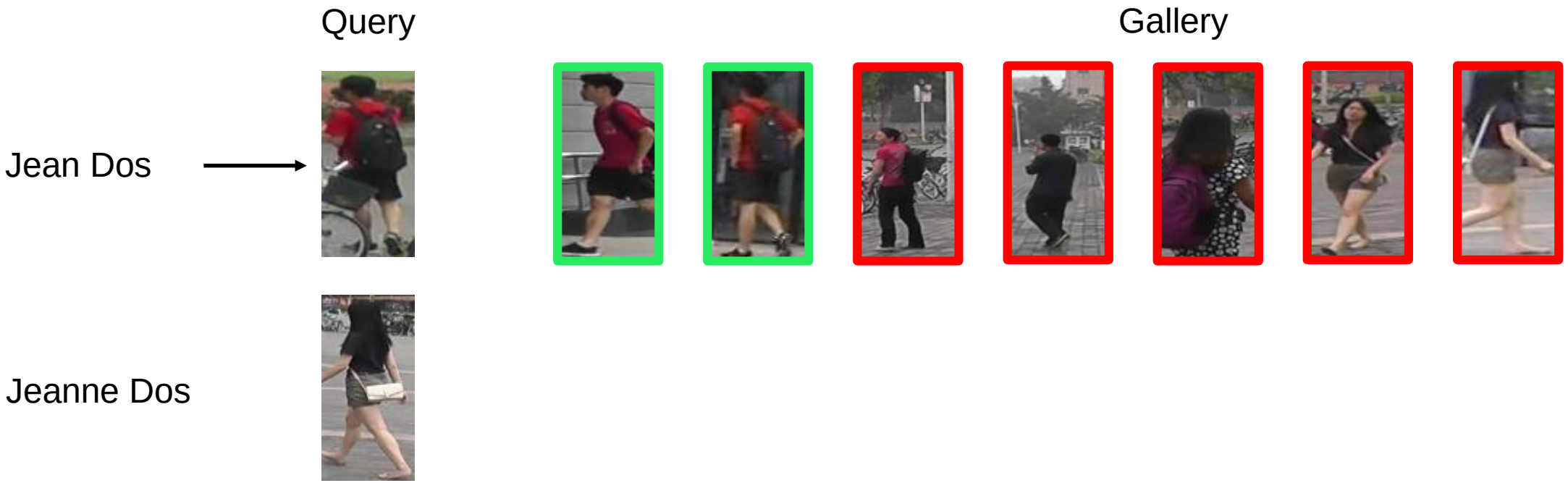
Query



Gallery

- **Open-set Ranking problem:**
  - Different classes between training and testing
  - Ranking a Gallery from most to least similar to a Query

PERSON RE-IDENTIFICATION

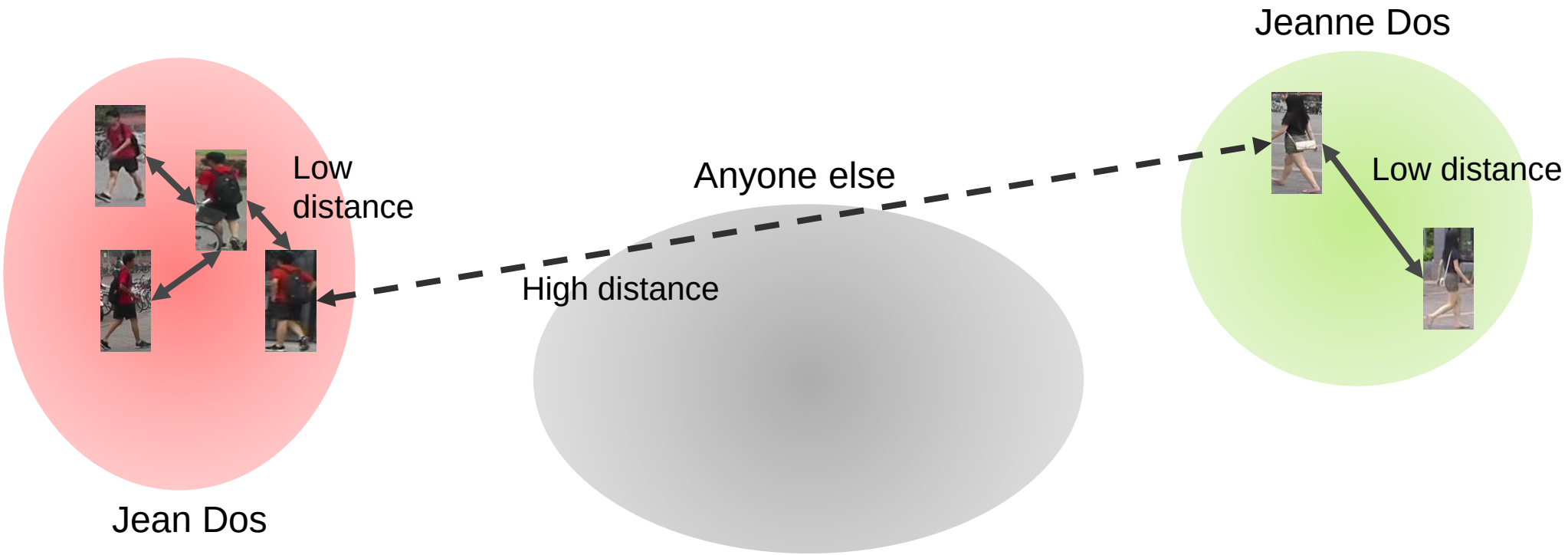


- **Open-set Ranking problem:**
  - Different classes between training and testing
  - Ranking a Gallery from most to least similar to a Query

PERSON RE-IDENTIFICATION

- **Open-set Ranking problem:**
  - Different classes between training and testing
  - Ranking a Gallery from most to least similar to a Query

METRIC LEARNING

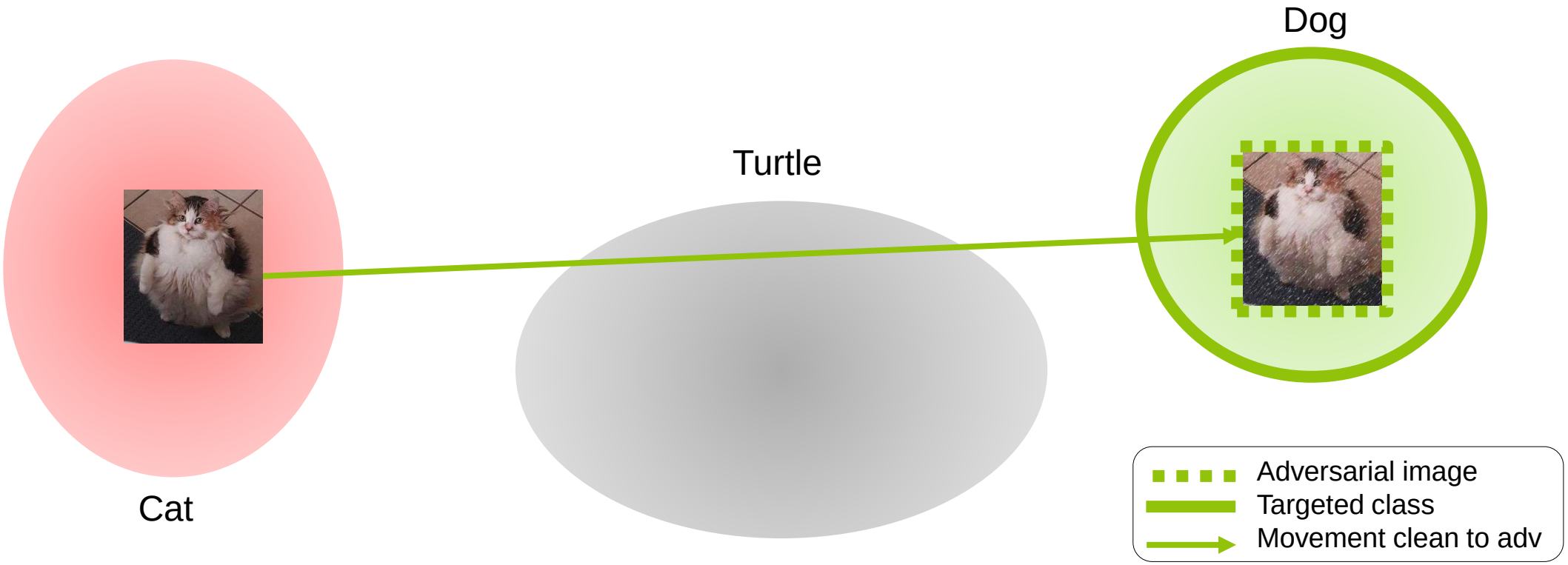


- **Low distance with same person**
- **High distance with different person**

METRIC ATTACKS

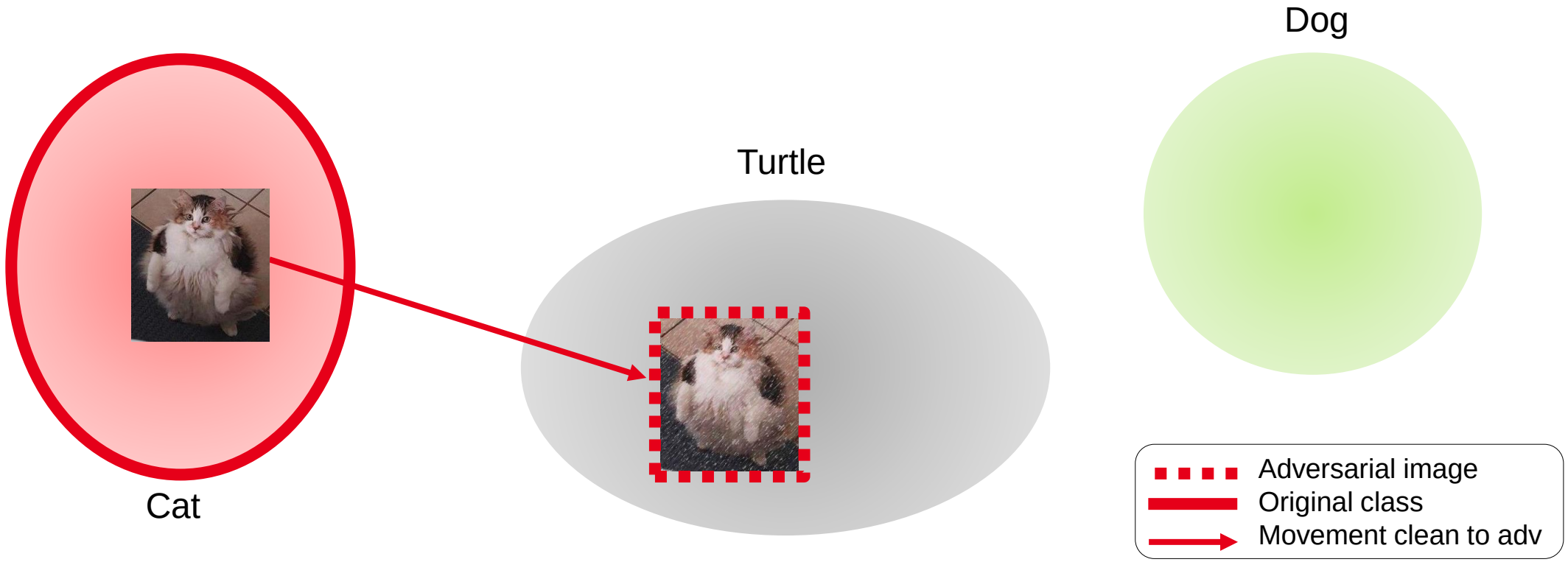
Quentin Bouniot, Romaric Audigier, Angélique Loesch

CLASSIFICATION ATTACKS: TARGETED ATTACKS



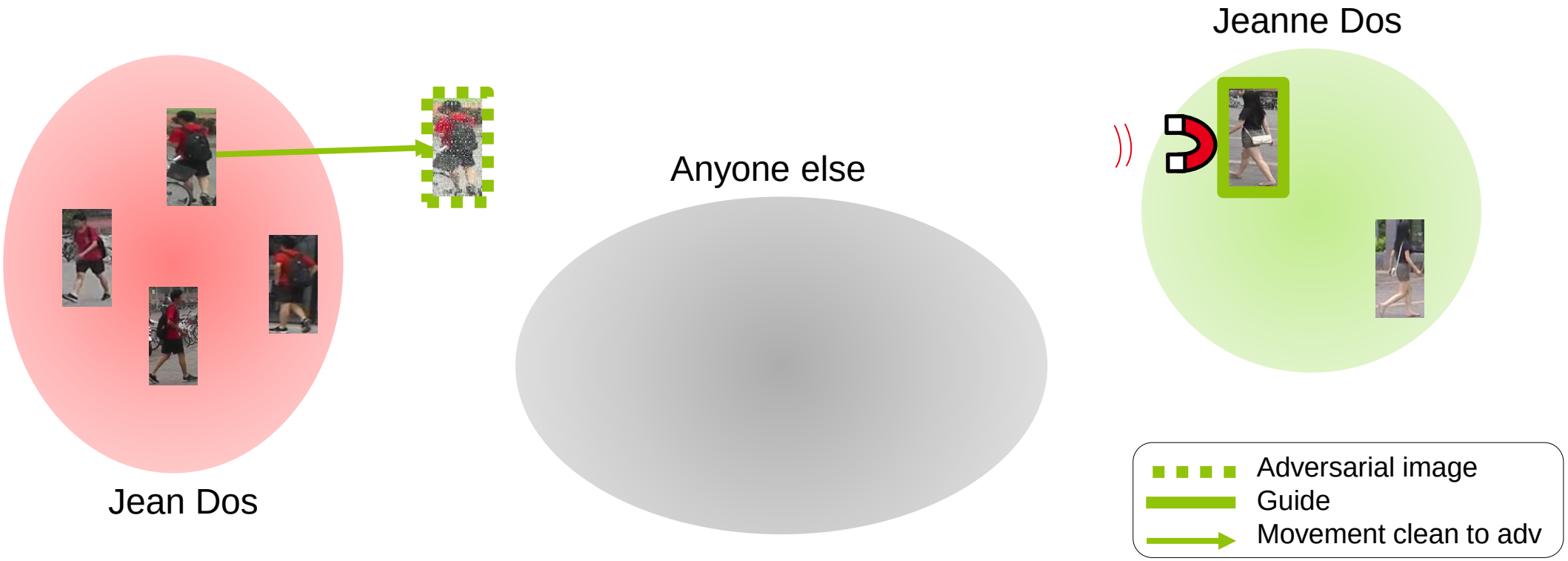
- **Using the class information:**
  - Attack models at the logit level
  - Keep class predictions as far away as possible from their proper class

CLASSIFICATION ATTACKS: NON-TARGETED ATTACKS



- **Using the class information:**
  - Attack models at the logit level
  - Keep class predictions as far away as possible from their proper class

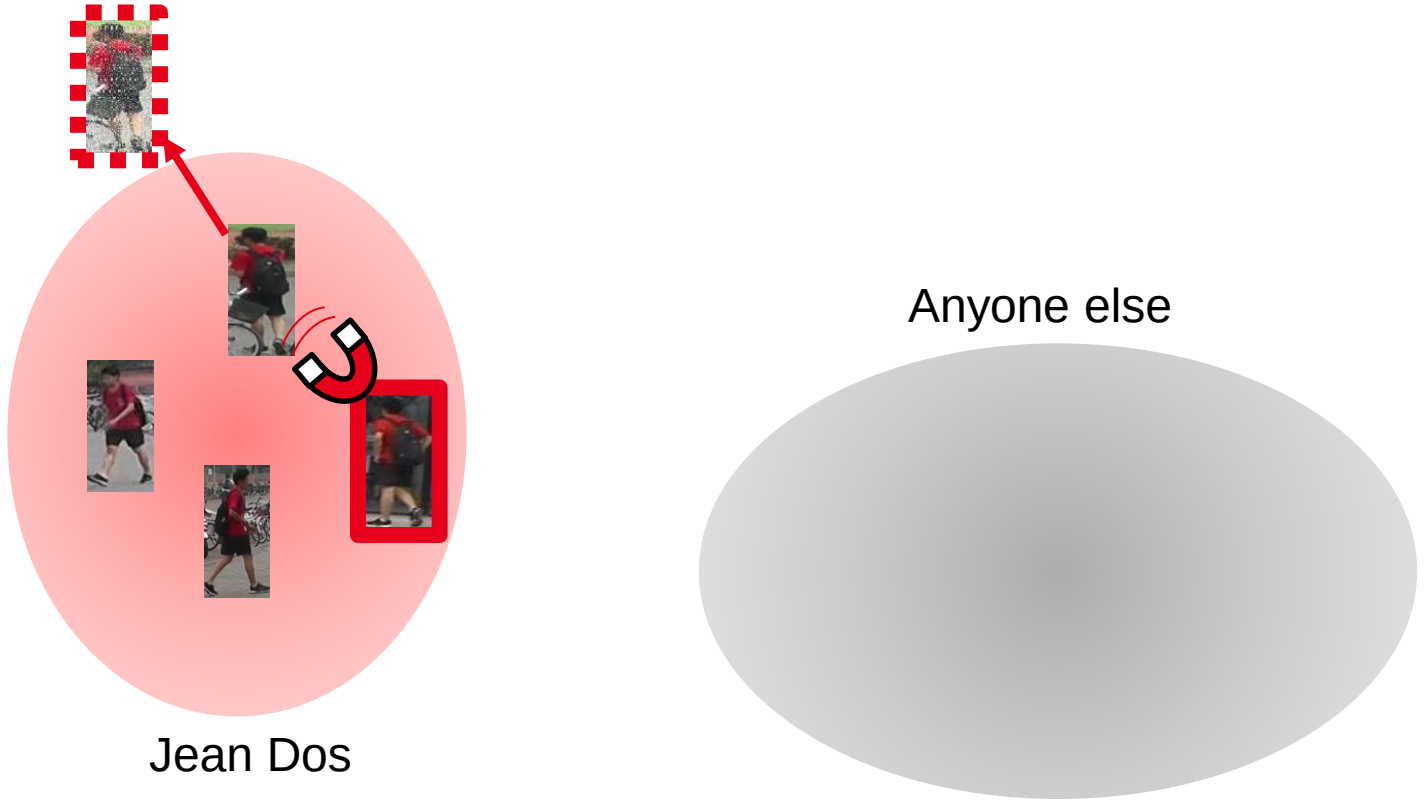
METRIC ATTACKS: PULLING GUIDE



- **As class information is not available in open-set: Attack the metric**
- **We need a *guide* ! (pushing / pulling)**
- **Existing attack: SG. FGSM/IFGSM/MIFGSM [2]**

[2]: Bai S. et al. 2019. *Metric Attack and Defense for Person Re-identification.* In arXiv.

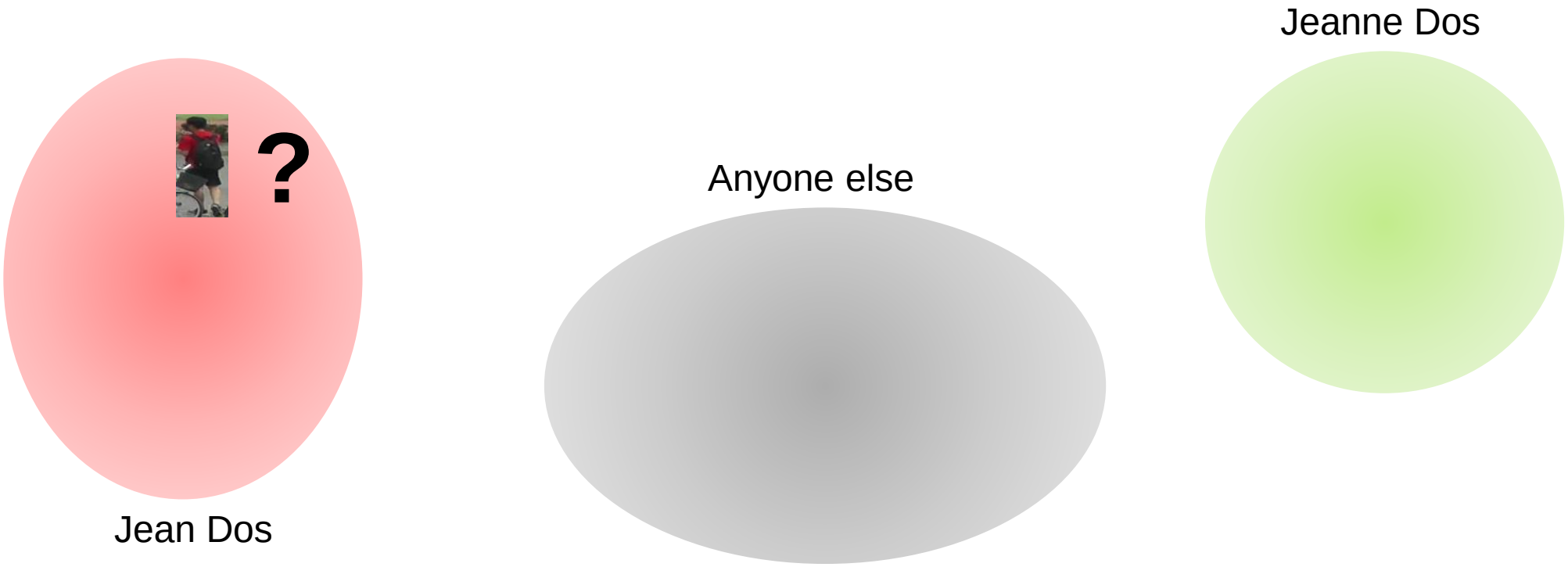
METRIC ATTACKS: PUSHING GUIDE



- **As class information is not available in open-set: Attack the metric**
- **We need a *guide* ! (pushing / pulling)**
- **Existing attack: SG. FGSM/IFGSM/MIFGSM [2]**

[2]: Bai S. et al. 2019. *Metric Attack and Defense for Person Re-identification.* In arXiv.

METRIC ATTACKS: ARTIFICIAL GUIDE



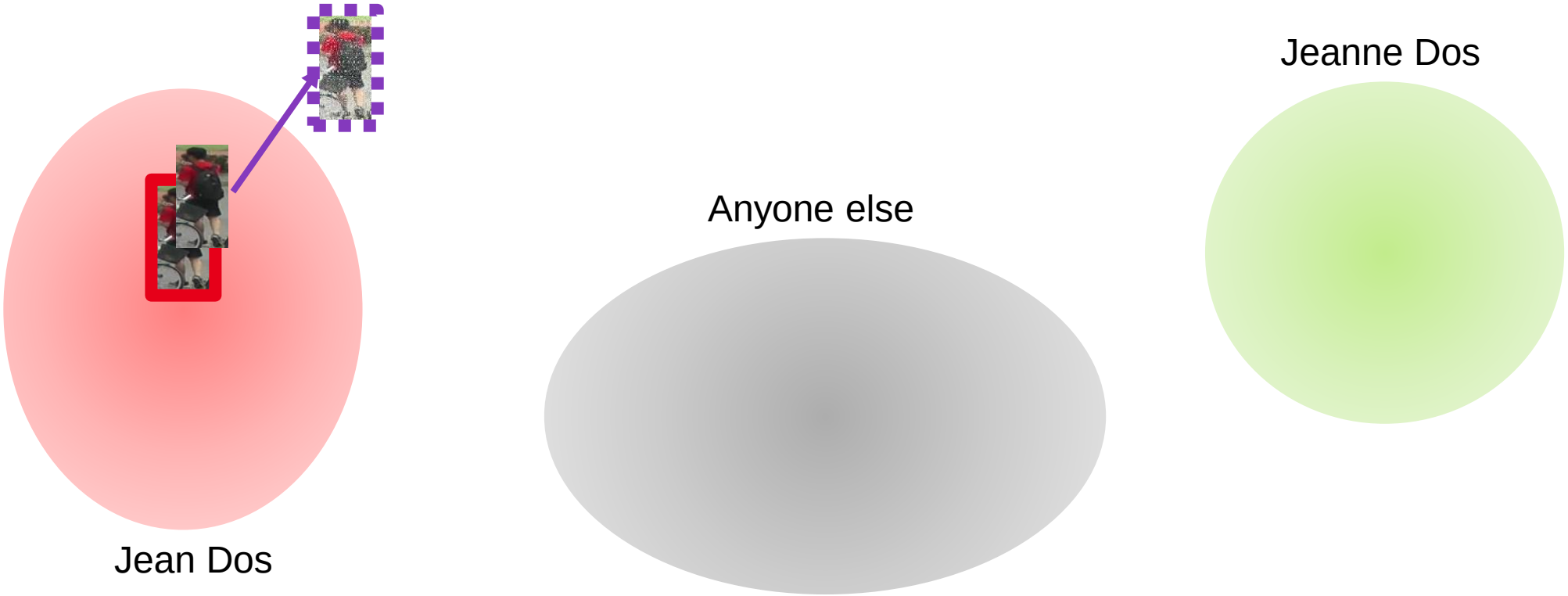
- **If we don’t have access to additional images: Where is the guide ?**
- **Construct an *artificial guide* from the image under attack**
- **Existing attack: ODFA [1]**

[1]: Zhedong Z. et al. 2018. *Open-set Adversarial Examples.* In arXiv.

OUR ATTACKS

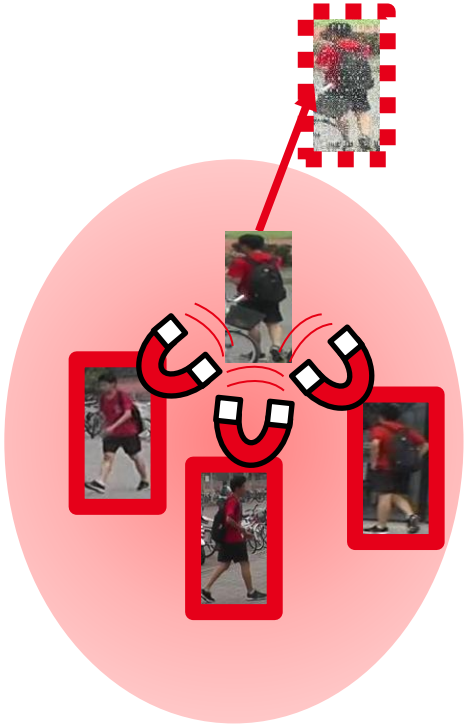
Quentin Bouniot, Romaric Audigier, Angélique Loesch

CONTRIBUTION:
SELF METRIC ATTACK (SMA)



- Use the *image under attack* as an ***artificial pushing guide***
- Move a *noisy copy* of the image away from the original image

CONTRIBUTION:
MULTIPLE PUSHING GUIDES



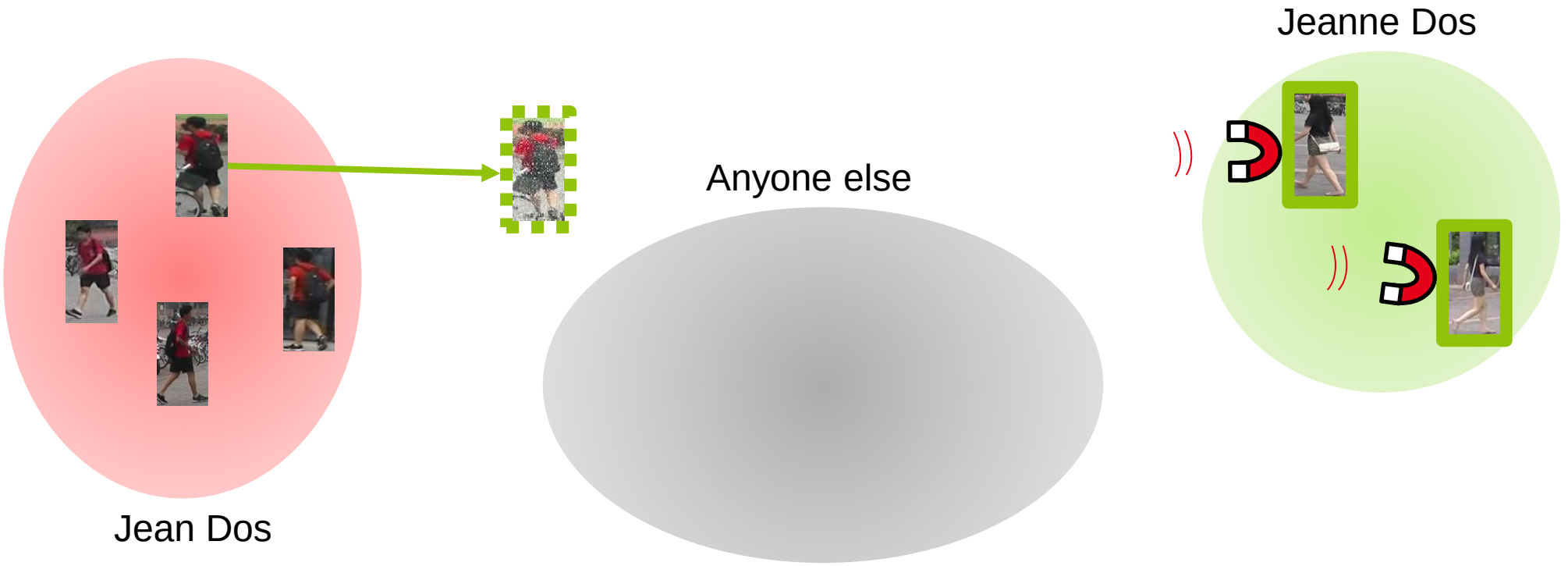
Jean Dos

Anyone else

Jeanne Dos

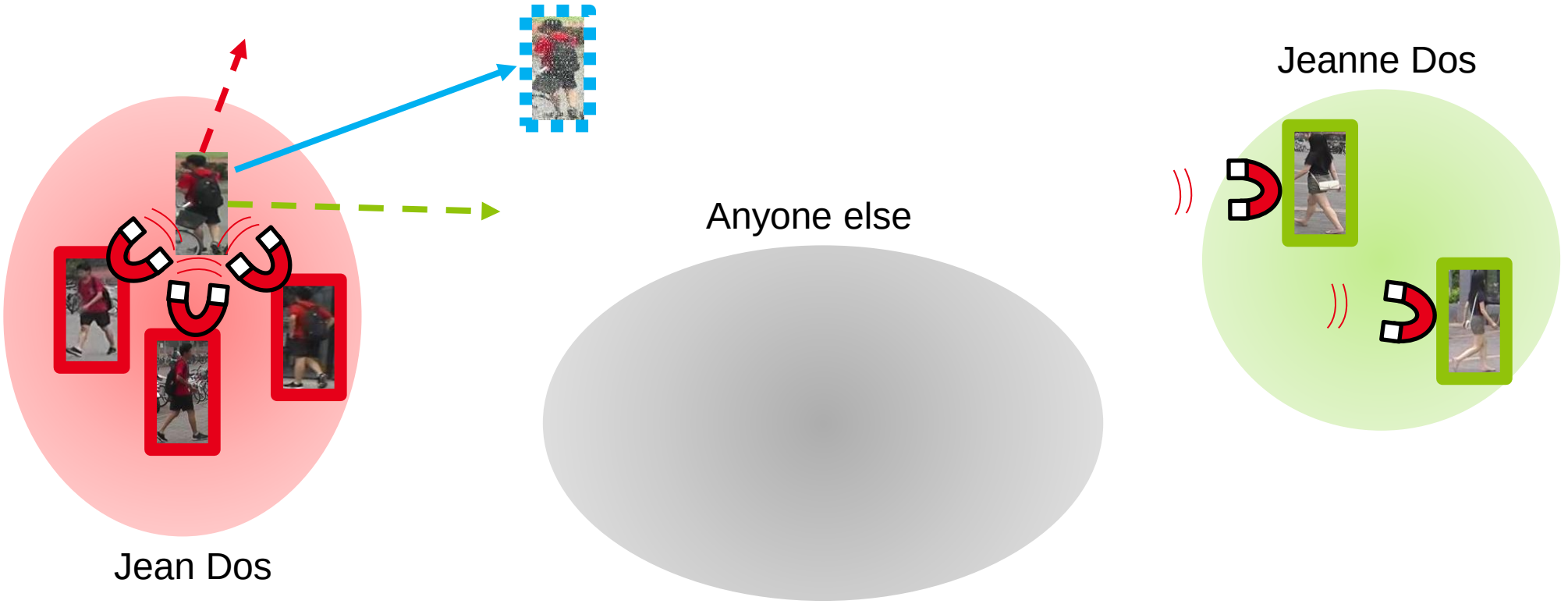
- **With multiple images, use multiple guides**
- **Better approximation of the best direction**

CONTRIBUTION:
MULTIPLE PULLING GUIDES



- With multiple images, use multiple guides
- Better approximation of the best direction

CONTRIBUTION:
FURTHEST-NEGATIVE ATTACK (FNA)



- **If we have a lot of images available:**
  - Combine multiple ***pushing guides*** and ***pulling guides*** from the ***furthest*** identity cluster
  - Make full use of the information (*images*) available

DEFENDING RE-IDENTIFICATION MODELS

Quentin Bouniot, Romaric Audigier, Angélique Loesch

GUIDE-SAMPLING ONLINE ADVERSARIAL TRAINING (GOAT)

Training Batch

- **Special care for *adversarial training [3]* with *metric attacks*:**
  - No guarantee that multiple images have the same identity in a training batch

[3]: Madry A. et al. 2017. *Towards Deep Learning Models Resistant to Adversarial Attacks.* In ICLR 2018.

GUIDE-SAMPLING ONLINE ADVERSARIAL TRAINING (GOAT)

- **Special sampling strategy:**
  - For each training image, sample pushing guides and pulling guides
- **Use the guides sampled to generate an adversarial batch**

CONCLUSION

- **Security and robustness are critical for Person Re-Identification**

- **Metric attacks require a guide:**
  - To increase the distance with the same identity (**pushing guide**)
  - To decrease the distance with another identity (**pulling guide**)

- **We proposed two metric attacks depending on availability of images:**
  - **Self Metric Attack (SMA):** strong self-sufficient attack
  - **Furthest-Negative Attack (FNA):** use all the information available

- **We improve robustness with GOAT:**
  - Extension of Adversarial Training [3] for an efficient defense against metric attacks

[3]: Madry A. et al. 2017. *Towards Deep Learning Models Resistant to Adversarial Attacks.* In ICLR 2018.

THANK YOU FOR LISTENING !
FEEL FREE TO ASK QUESTIONS

MORE INFO:

Paper:

Q. Bouniot, R. Audigier and A. Loesch,
« Vulnerability of Person Re-Identification Models to Metric Adversarial Attacks »,
2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),
Seattle, WA, USA, 2020, pp. 3450-3459,
doi: 10.1109/CVPRW50498.2020.00405.

Technical Blogpost:

https://qbouniot.github.io/article/2020/05/06/adv_reid.html

Contact:

quentin.bouniot@cea.fr

@QBouniot

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA SACLAY NANO-INNOV | BAT. 861 – PC142
91191 Gif-sur-Yvette Cedex - FRANCE
www-list.cea.fr