Towards a General Video-based Keystroke Inference Attack

Zhuolin Yang*, Yuxin Chen*, Zain Sarwar, Hadleigh Schwartz,
Ben Y. Zhao, Heather Zheng

*denotes equal contribution

SAND Lab
security, algorithms, networks and data

THE UNIVERSITY OF CHICAGO

Remote Work

We work and type in public, e.g., lounges, cafés, airports.

Remote Work

We work and type in public,
e.g., lounges, cafés, airports.

Sensitive text:

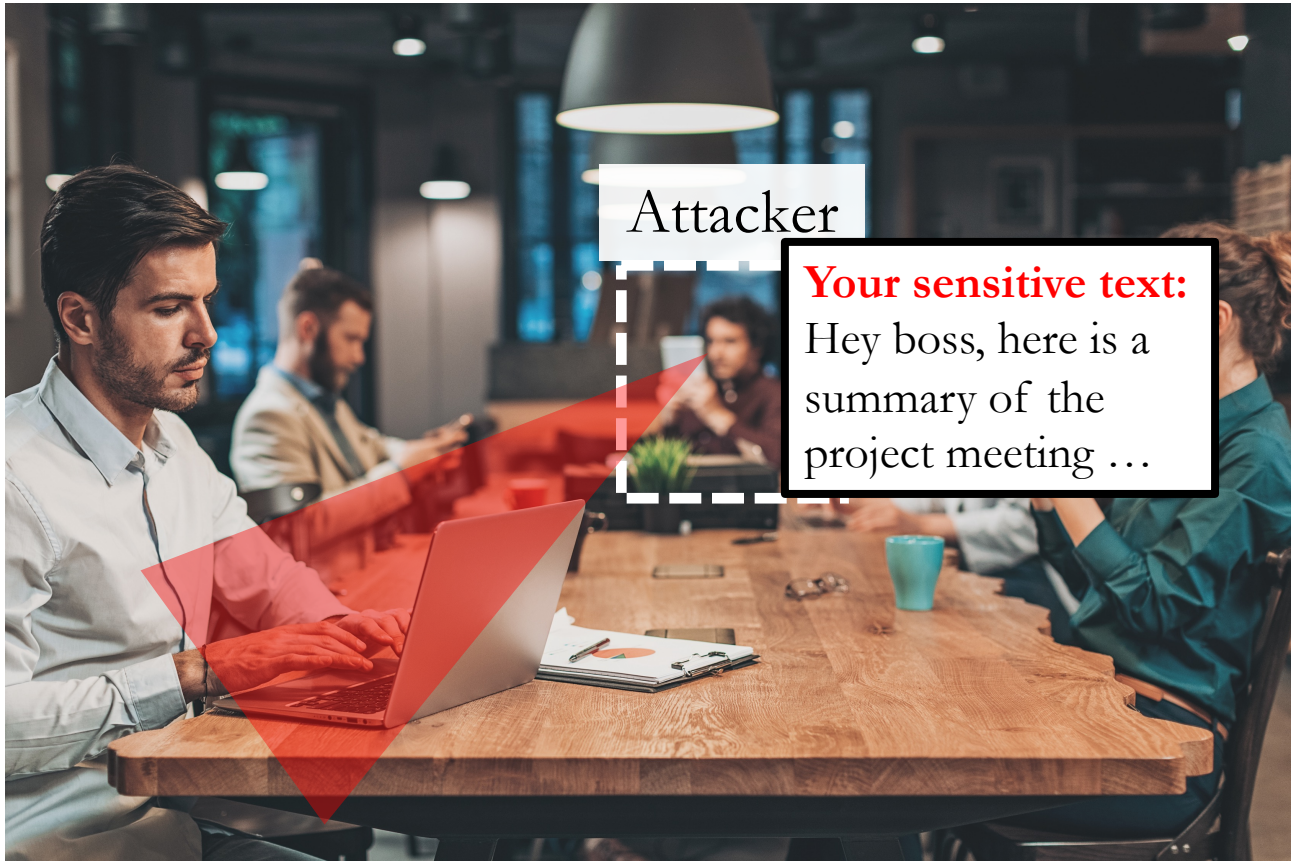
Work emails,
Research proposals,
Private documents,
……

Keystroke Inference Attack

Keystroke Inference Attack

Keystroke Inference Attack

Keystroke Inference Attack

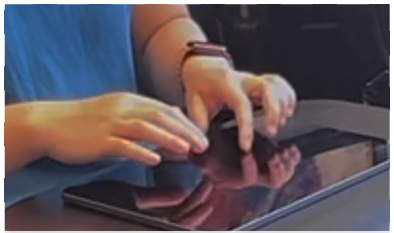


Keystroke Inference Attack

Practical Keystroke Inference Attack

Related works assume external information

- Labeled data of the target
- Keyboard size and layout
- Finger reflection
- Very close sensors (cm)



Finger reflection

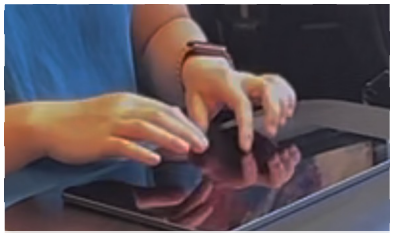
Sensor under the target's table

Practical Keystroke Inference Attack

Related works assume external information

- Labeled data of the target
- Keyboard size and layout
- Finger reflection
- Vary-to-vary sensors (vib)
- .....

No external information



Finger reflection

Sensor under the target's table

Threat Model

A general attack:
No external information

1. **The target:** types in English (10+ mins)

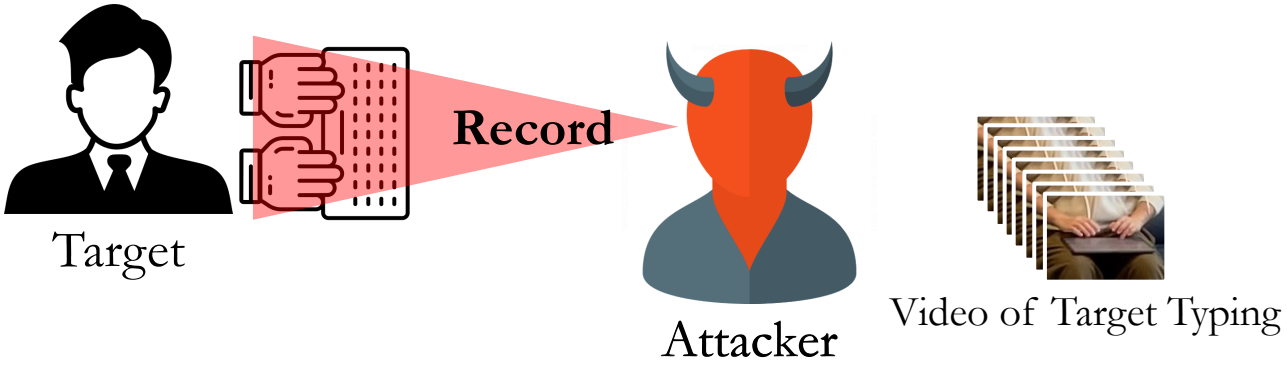
Threat Model

A general attack:
No external information

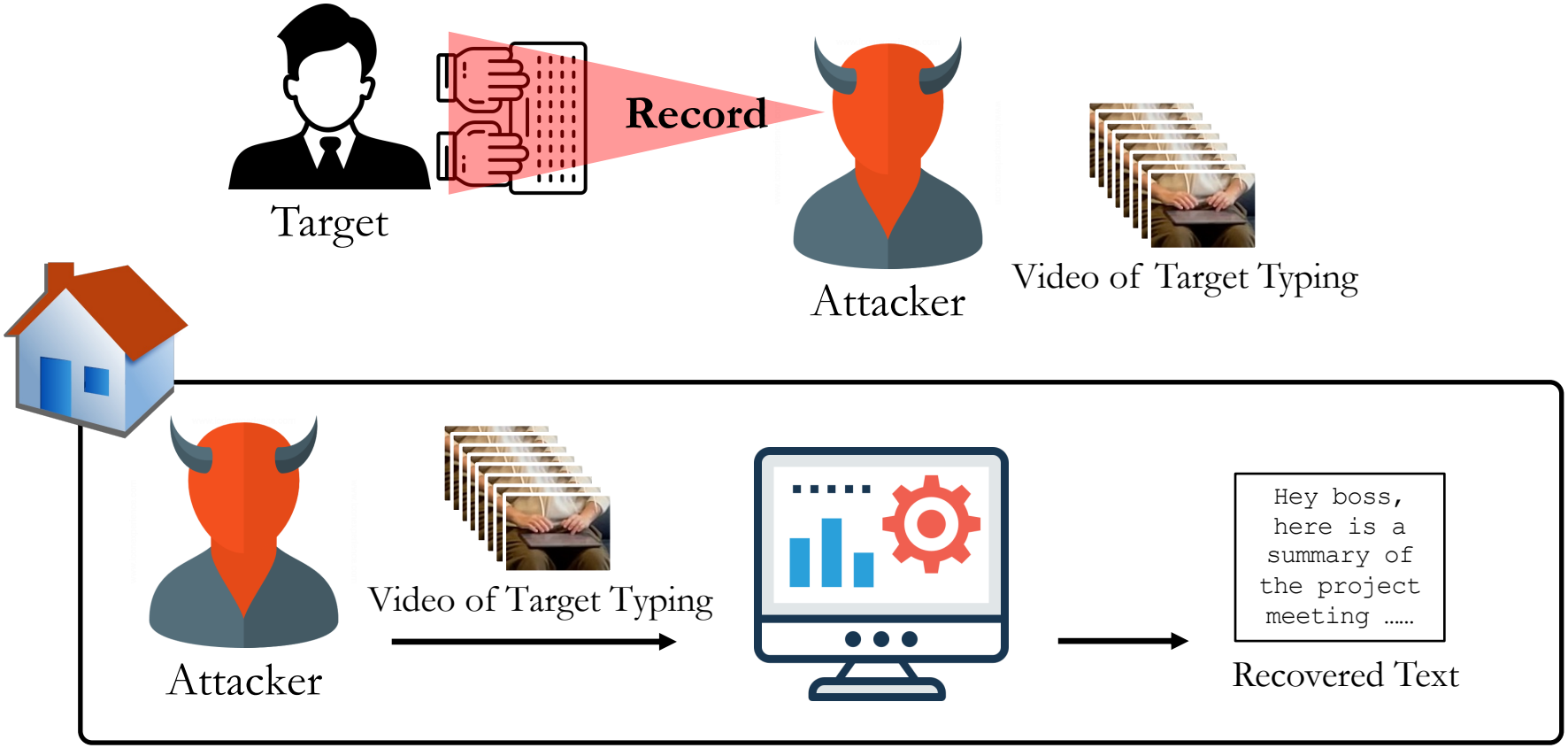
1. **The target:** types in English (10+ mins)

2. **The attacker:** 1 RGB camera
   - Unobstructive view of hands

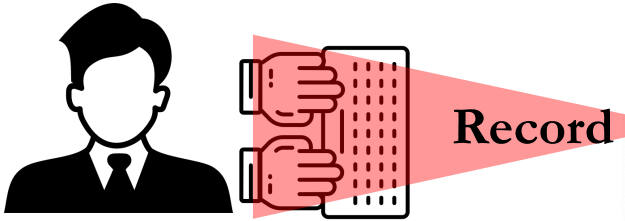
Attack Overview

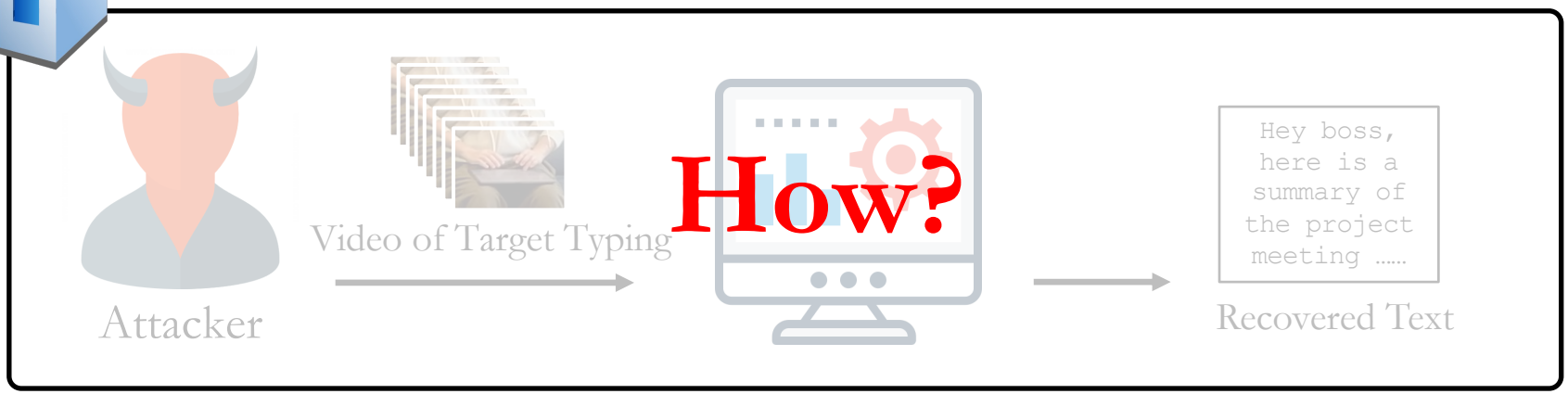


Attack Overview

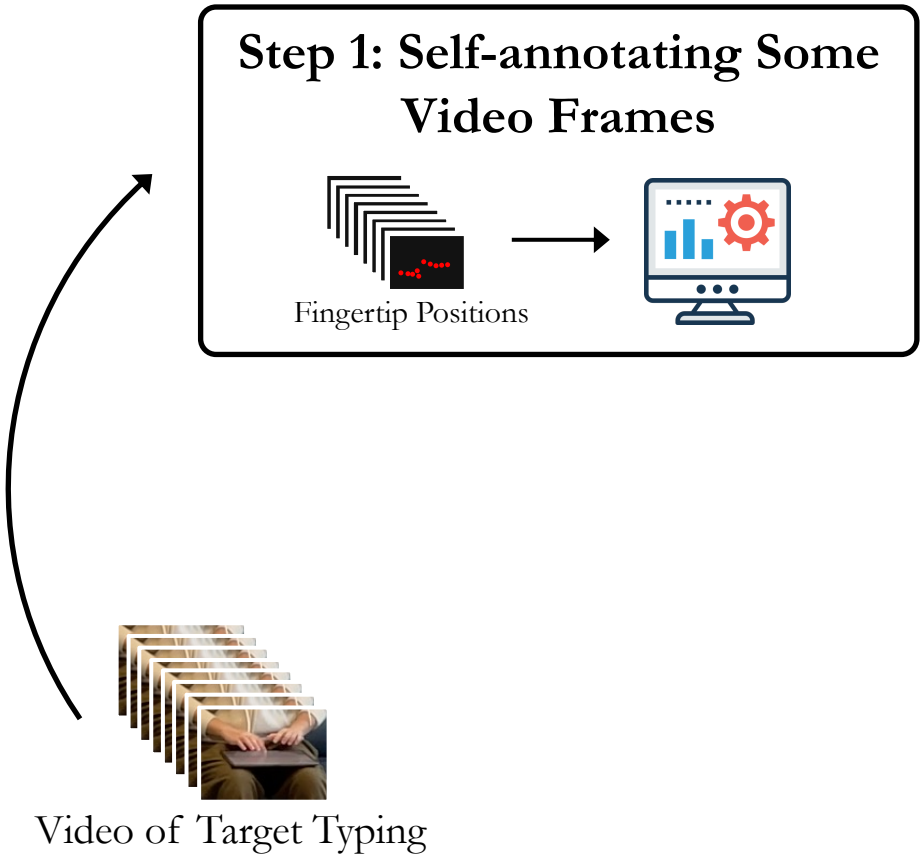


Attack Overview

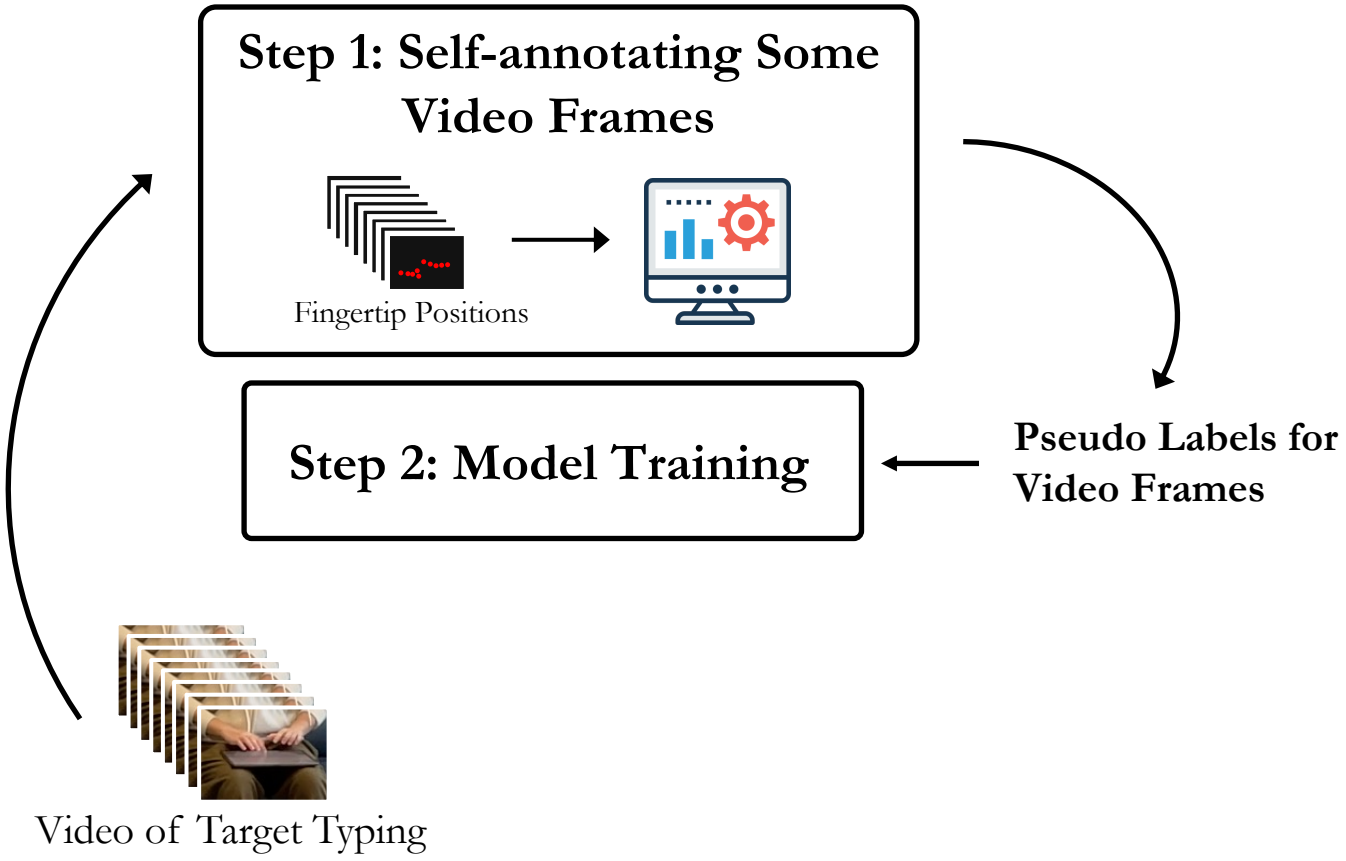




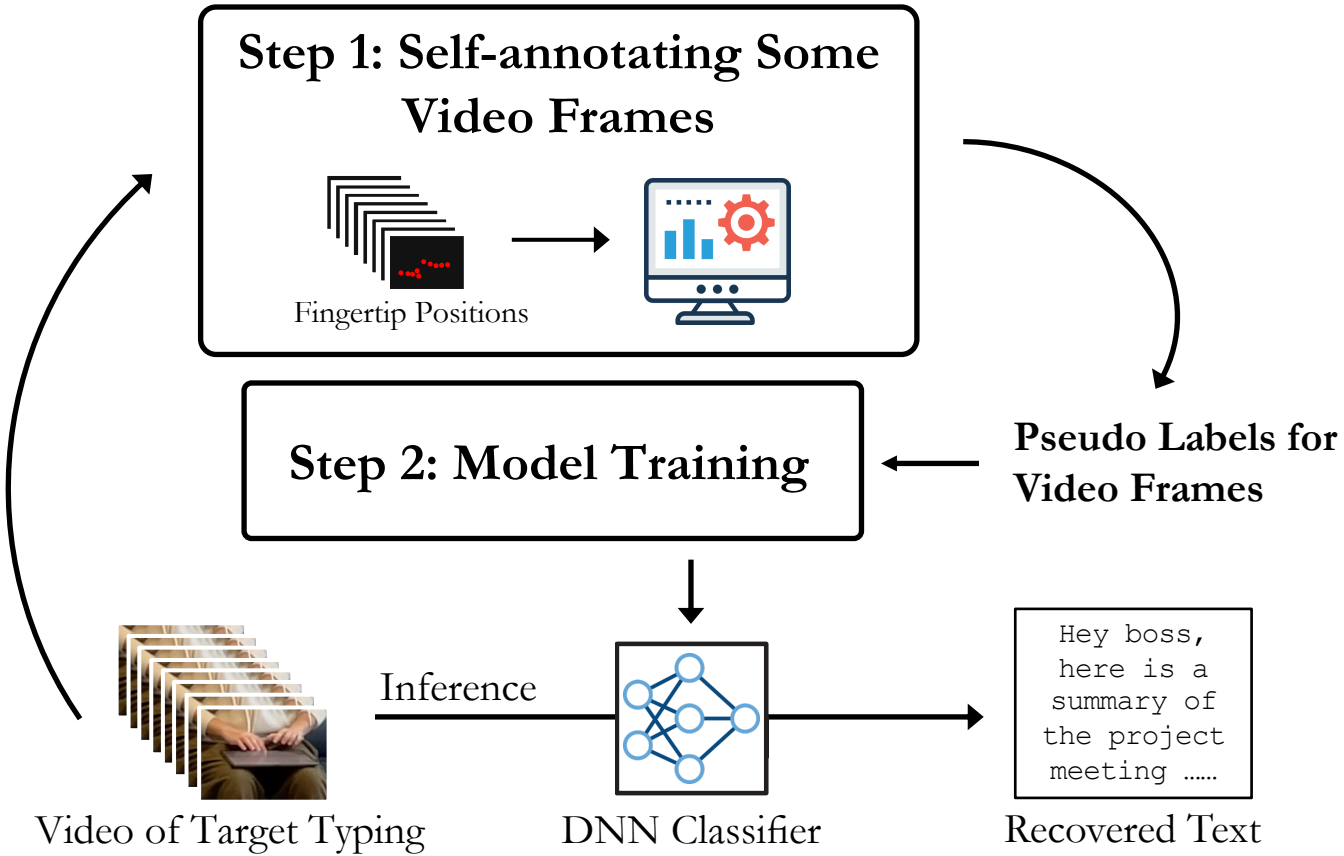
Design Intuition: Self-Supervised Learning



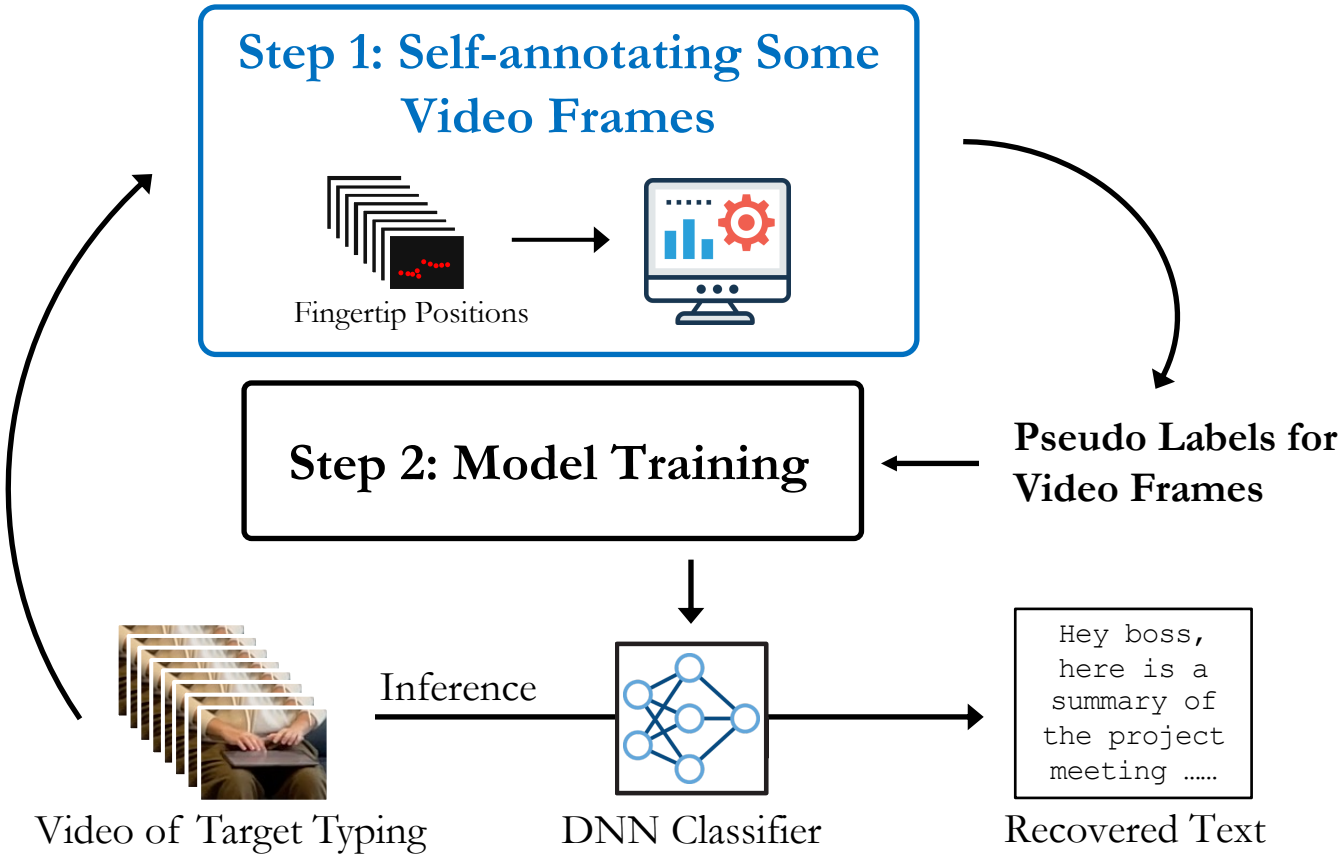
Design Intuition: Self-Supervised Learning



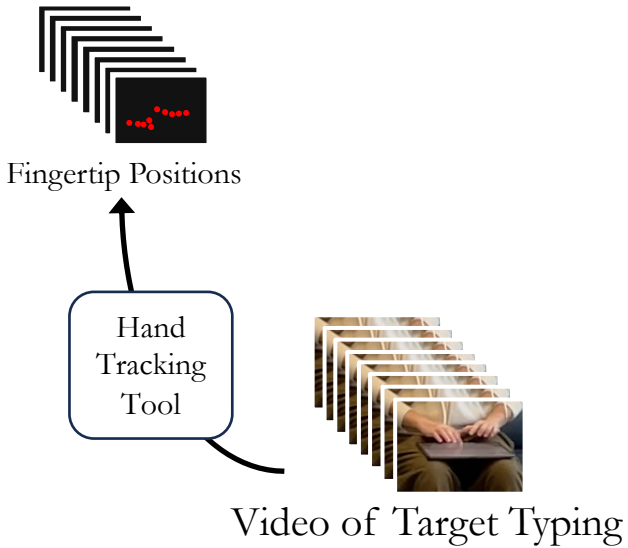
Design Intuition: Self-Supervised Learning



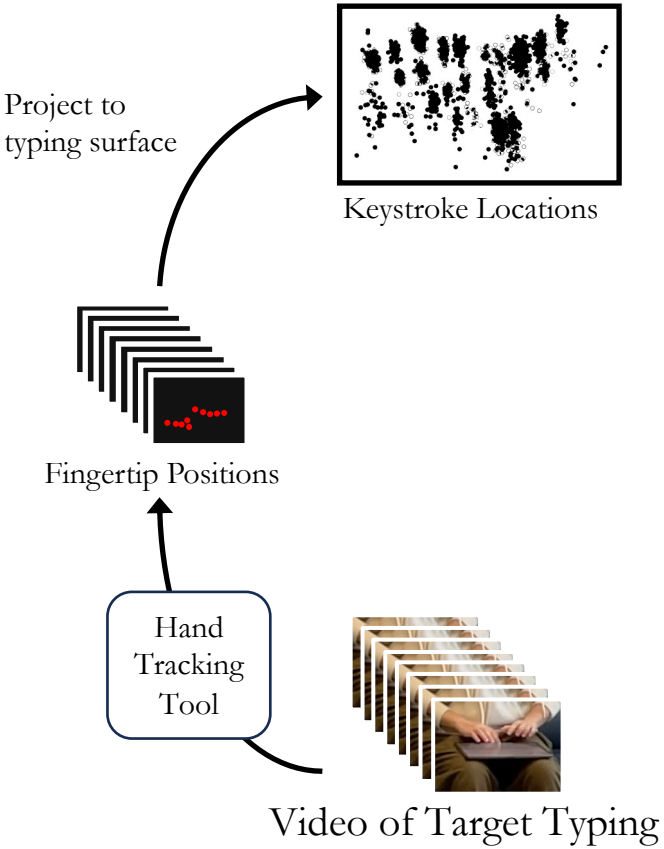
Design Intuition: Self-Supervised Learning



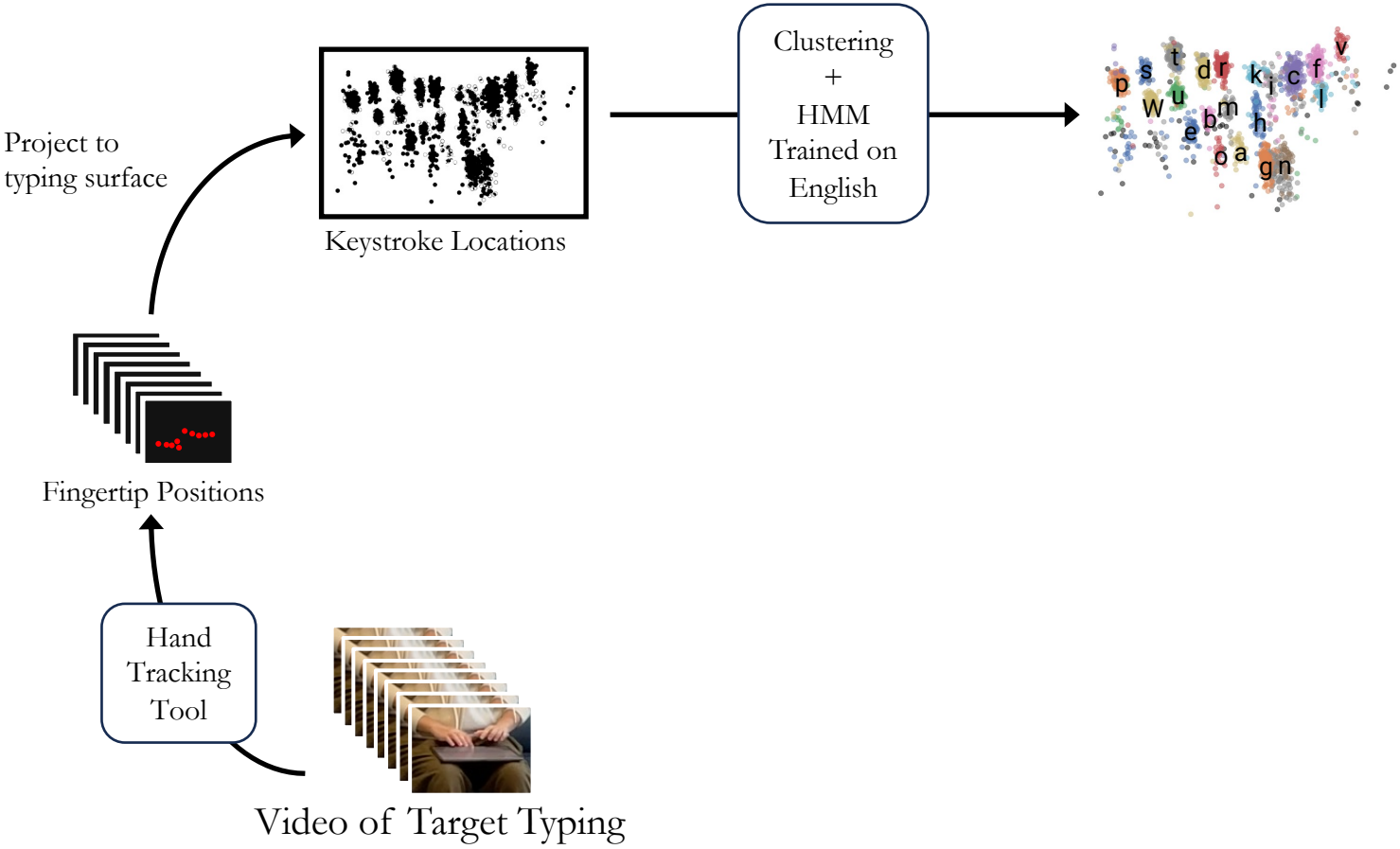
Step 1 – Self-annotating Some Video Frames



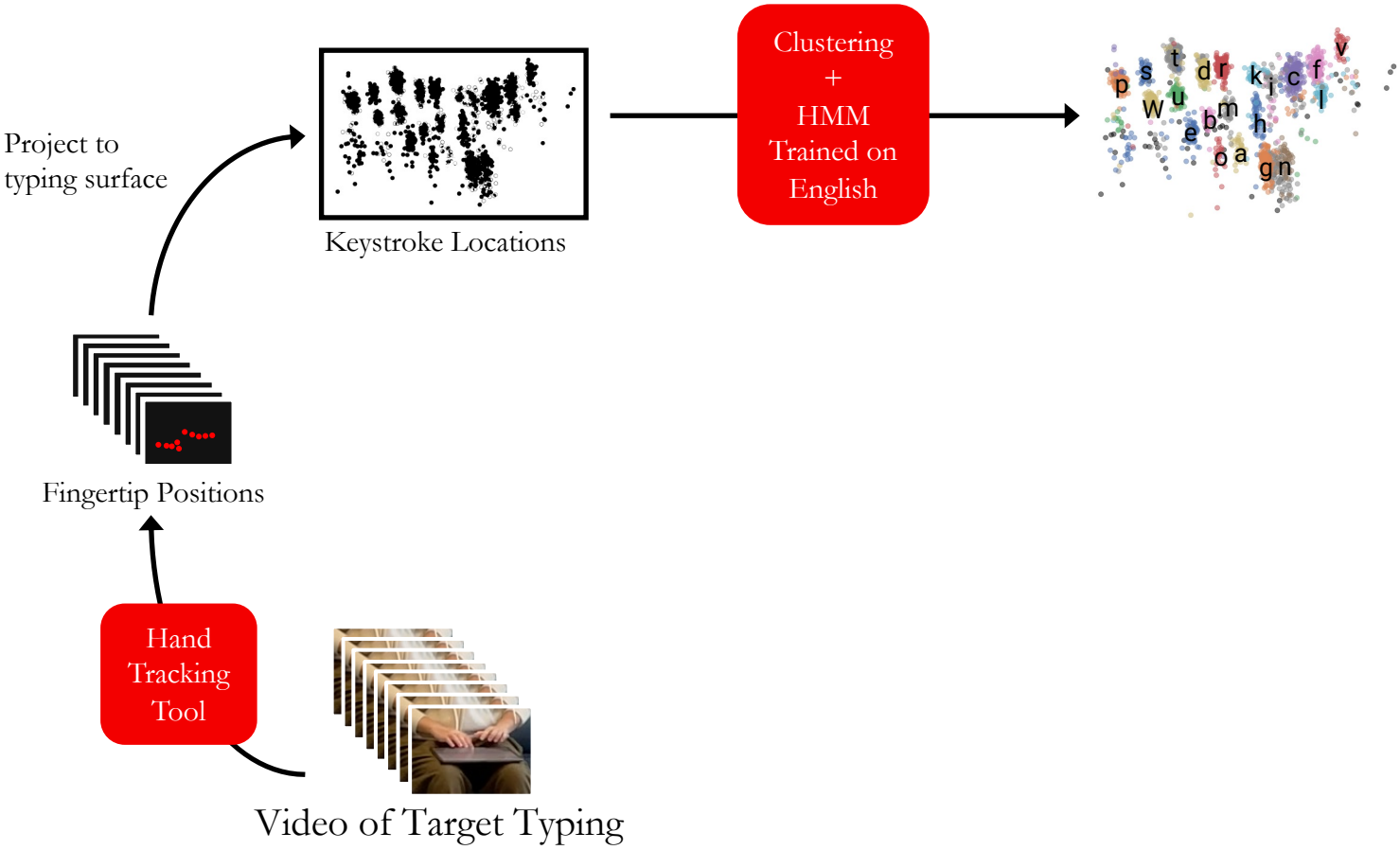
Step 1 – Self-annotating Some Video Frames



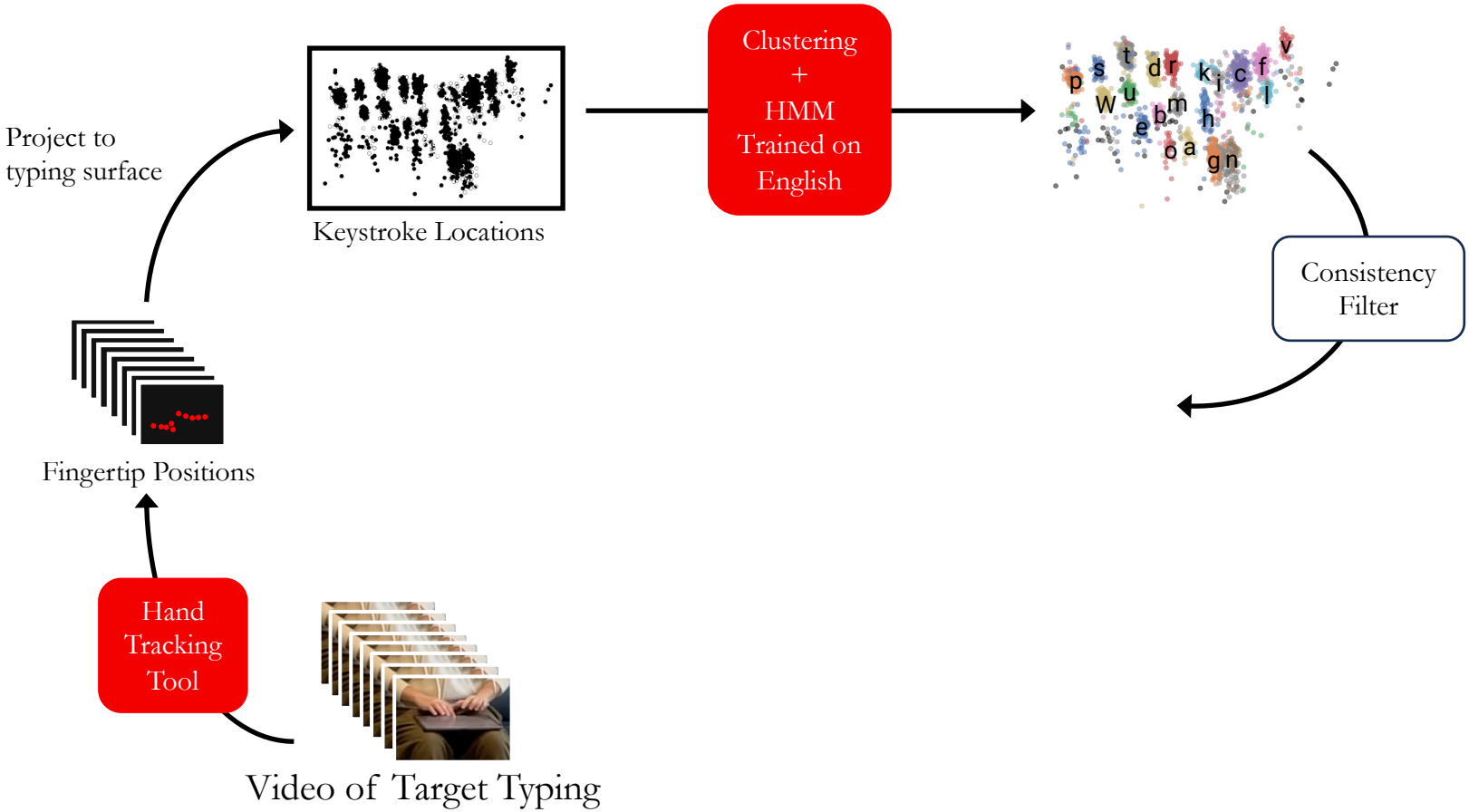
Step 1 – Self-annotating Some Video Frames



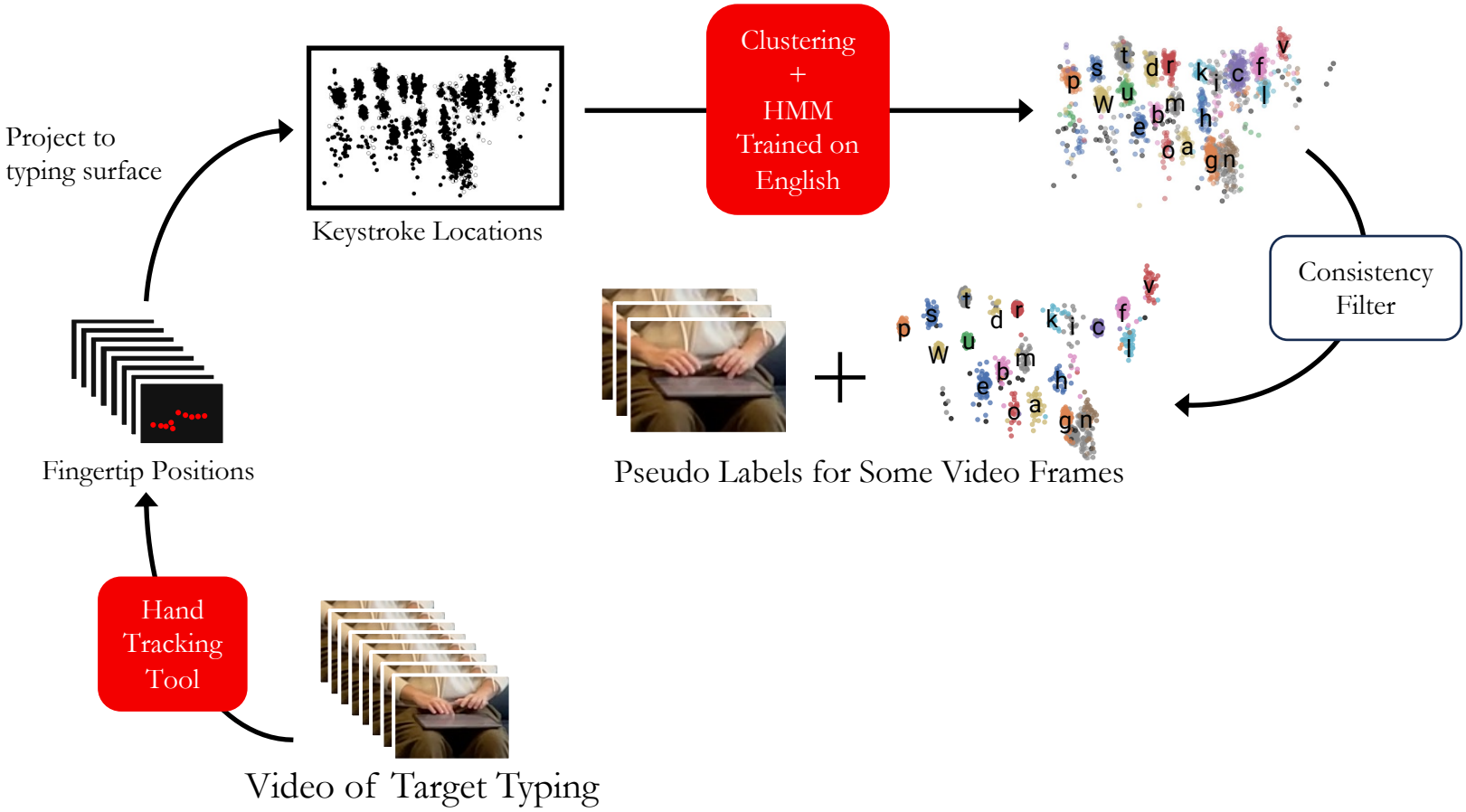
Step 1 – Self-annotating Some Video Frames



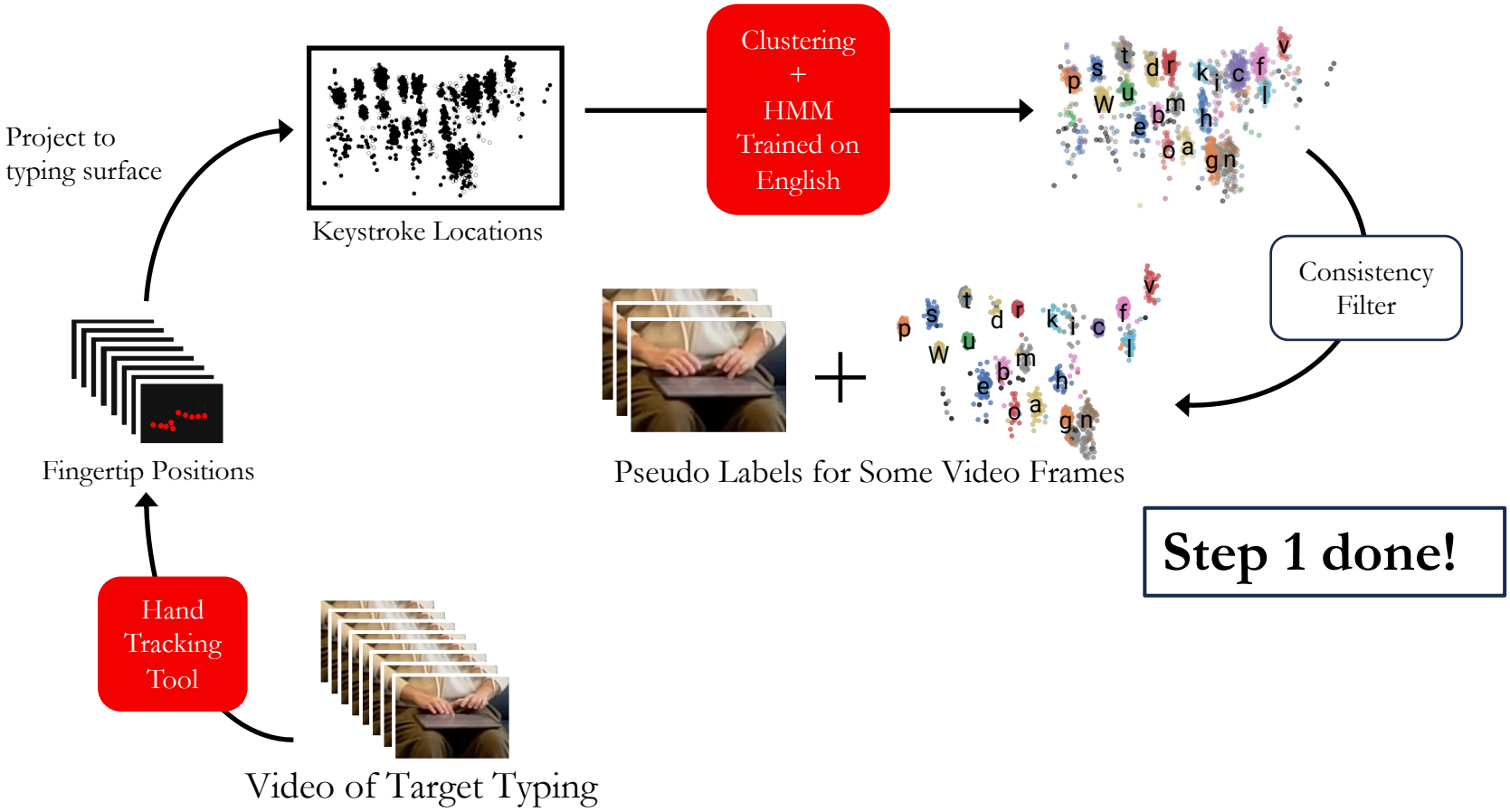
Step 1 – Self-annotating Some Video Frames



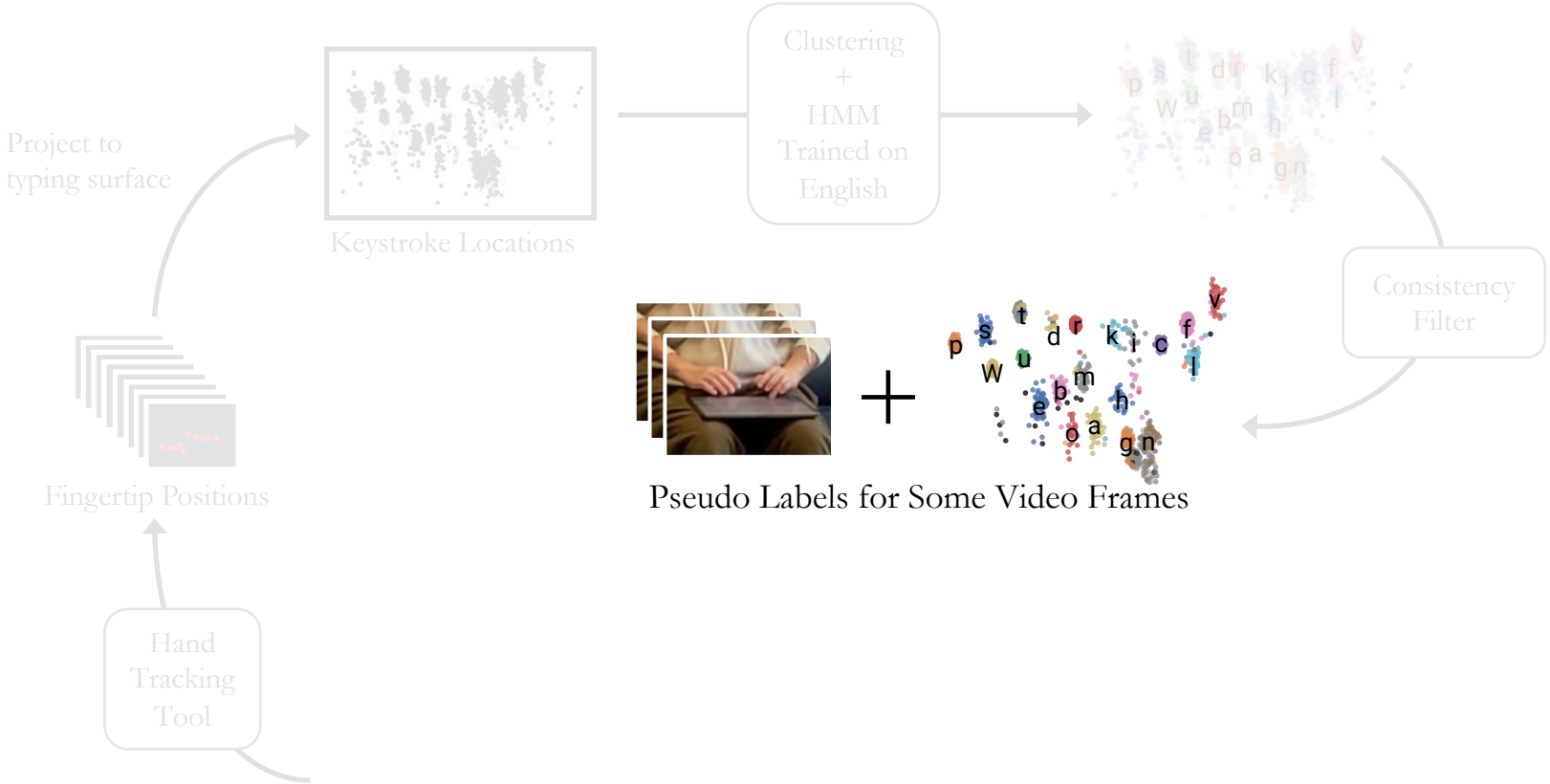
Step 1 – Self-annotating Some Video Frames



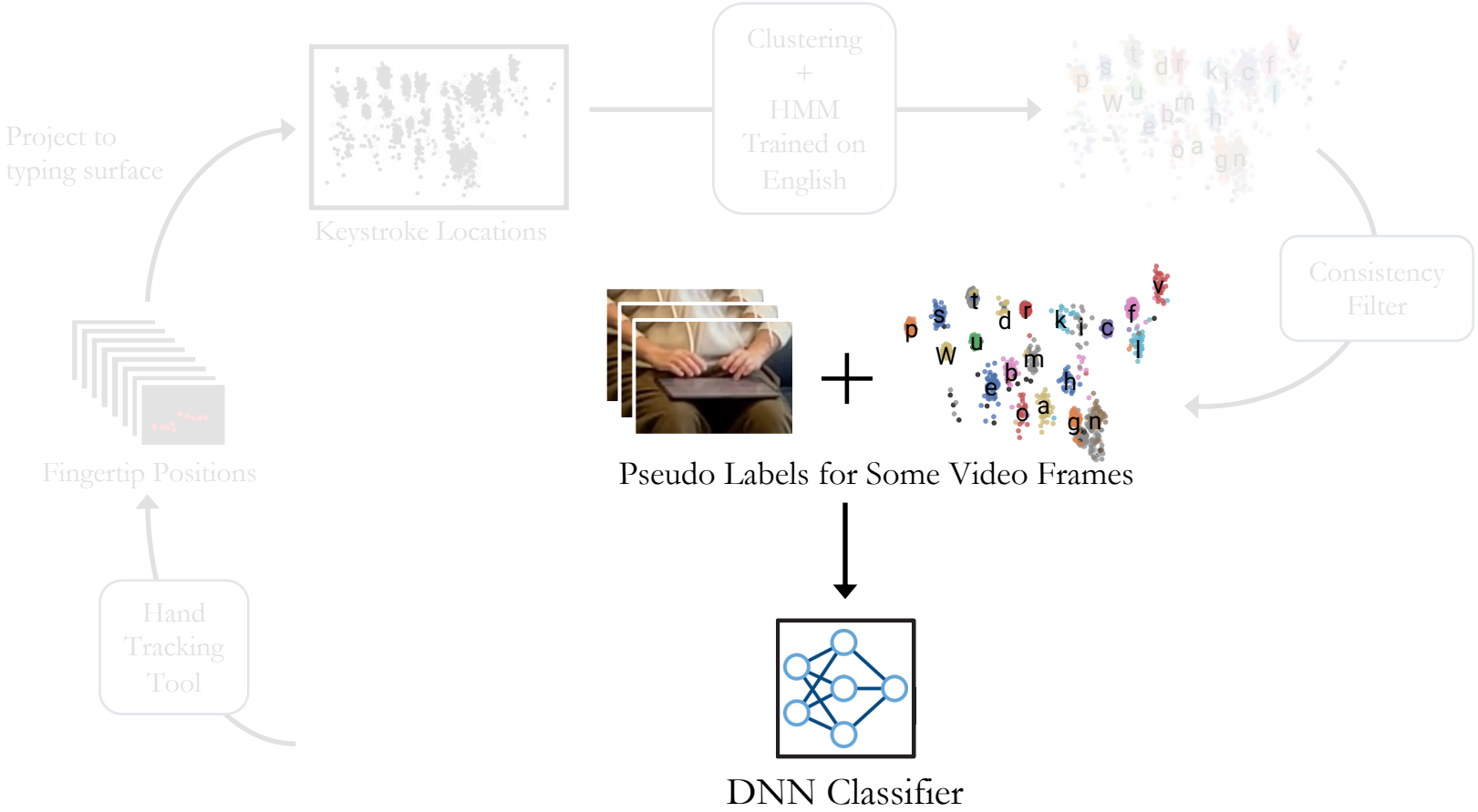
Step 1 – Self-annotating Some Video Frames



Step 2 – Model Training



Step 2 – Model Training



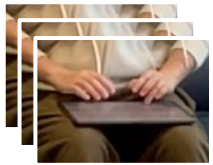
Step 2 – Model Training

Project to typing surface

Keystroke Locations

Clustering + HMM Trained on English

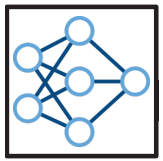
Consistency Filter



+

Pseudo Labels for Some Video Frames

Noise-Aware Training

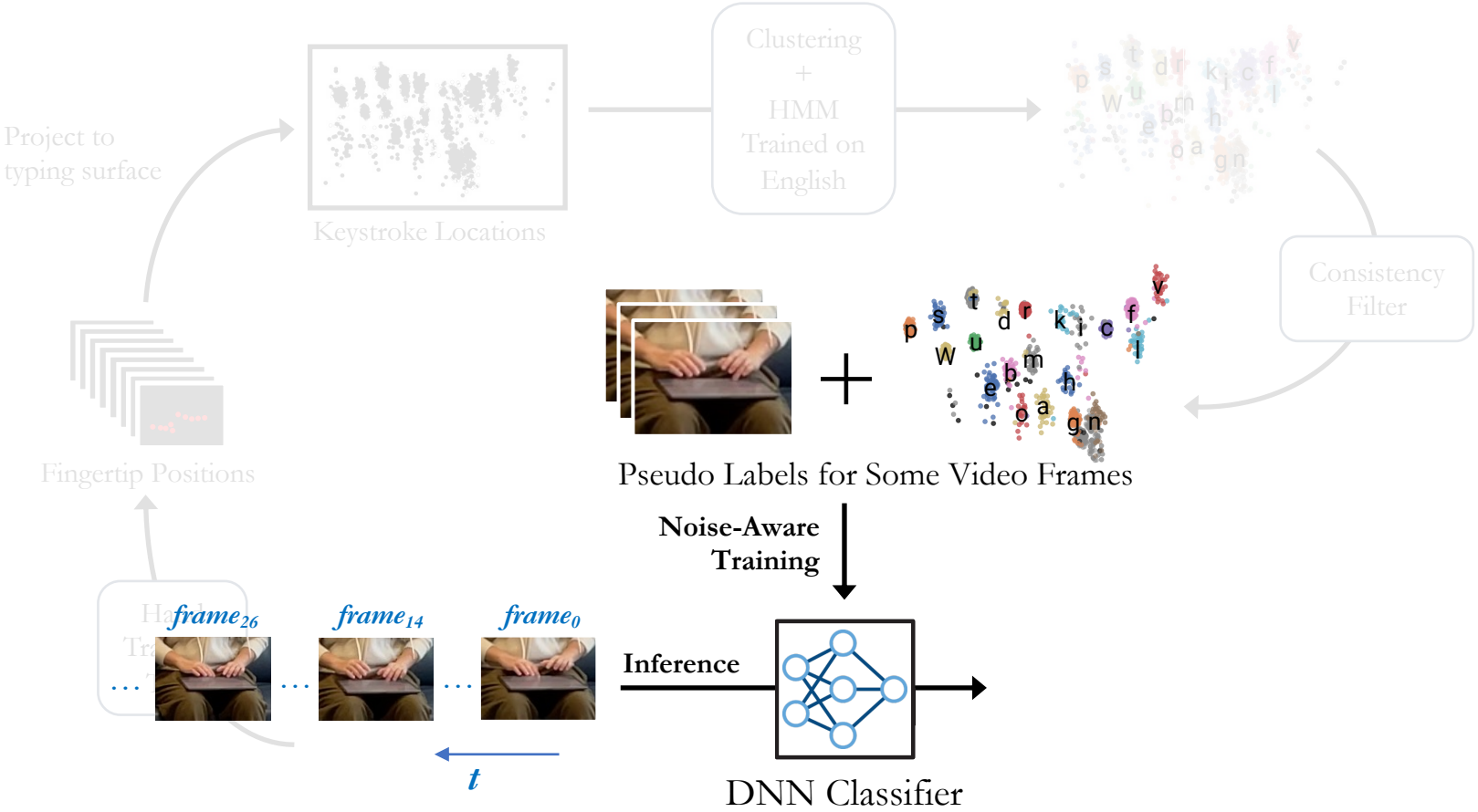


DNN Classifier

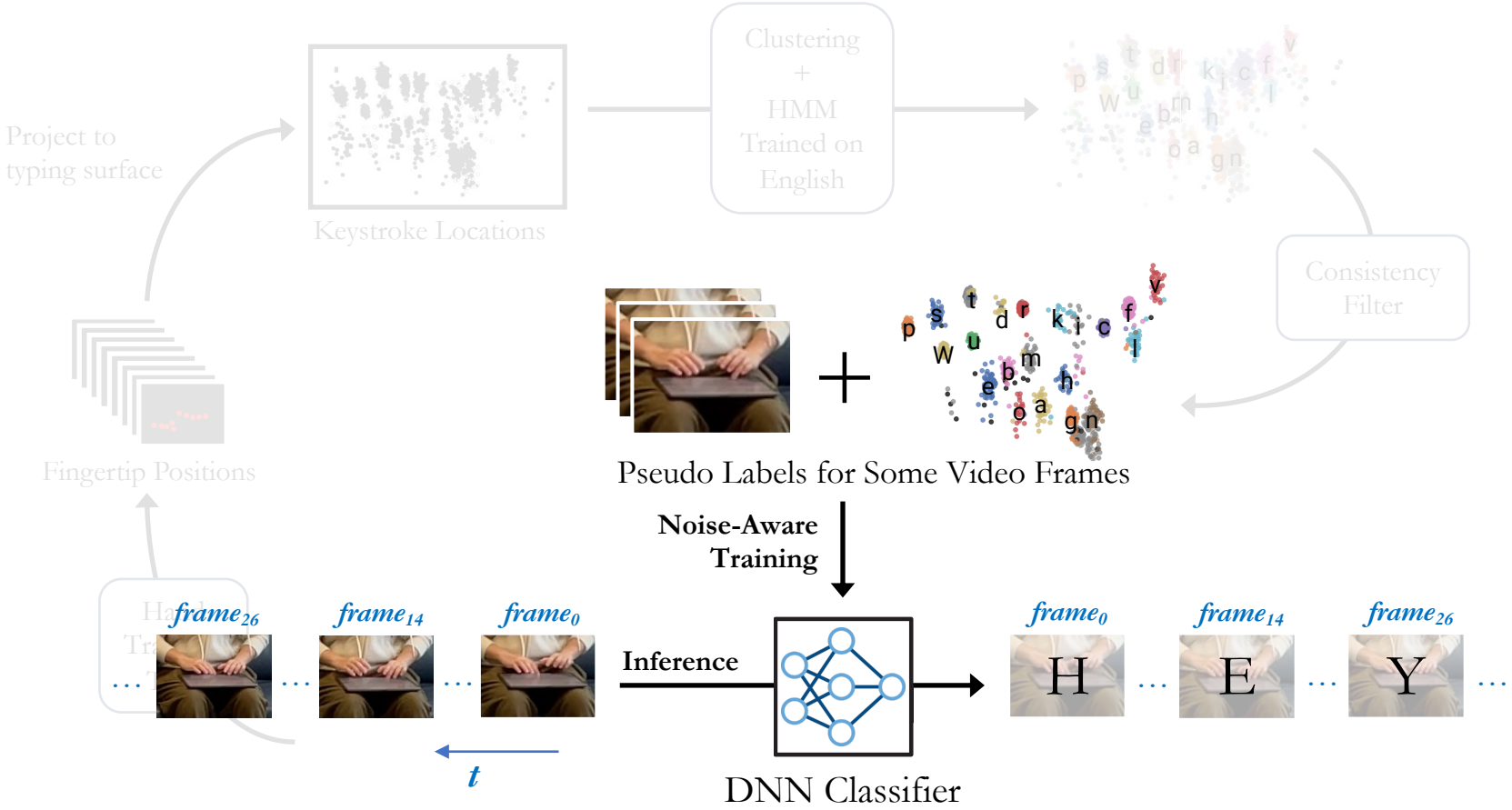
Fingertip Positions

Hand Tracking Tool

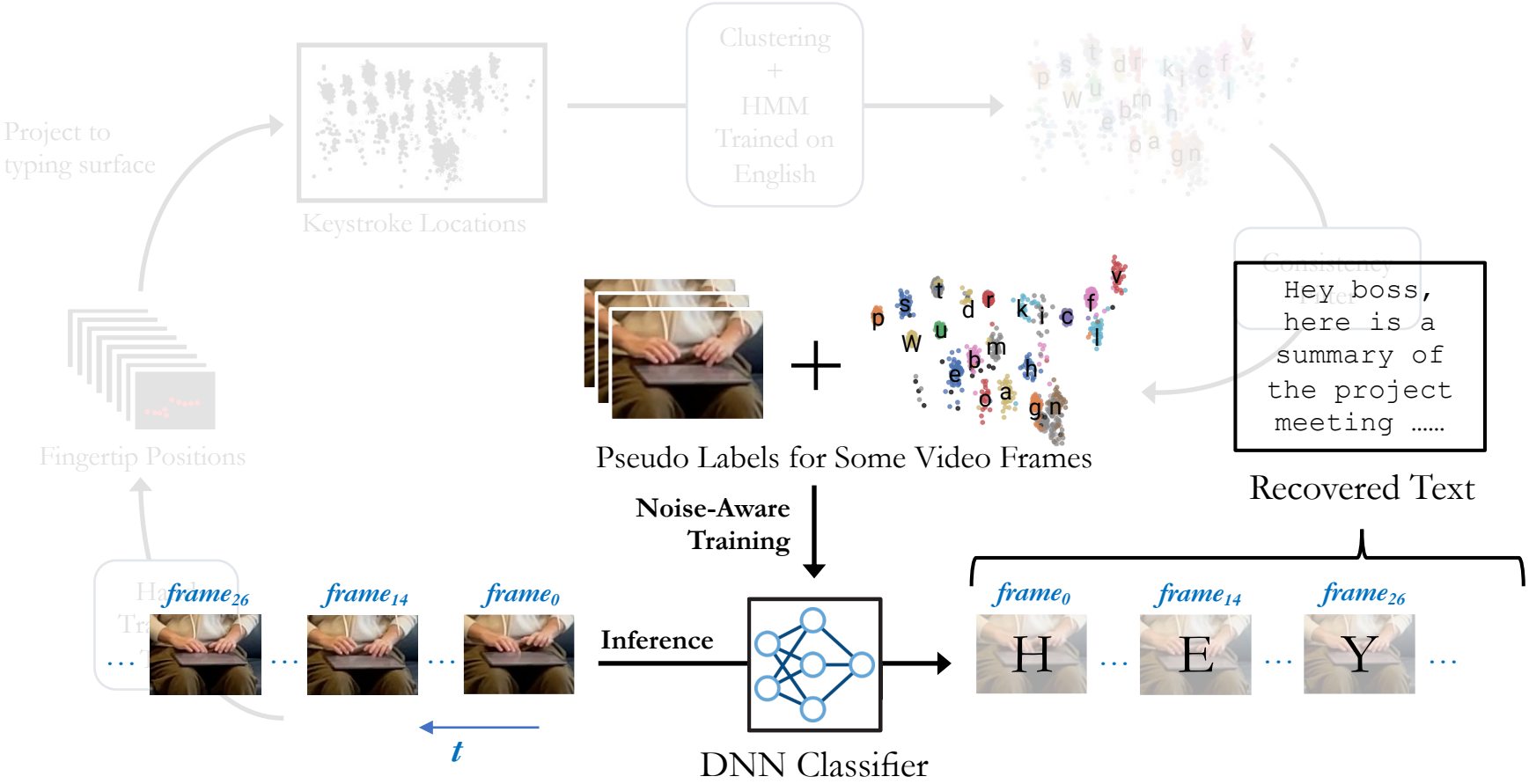
Step 2 – Model Training



Step 2 – Model Training

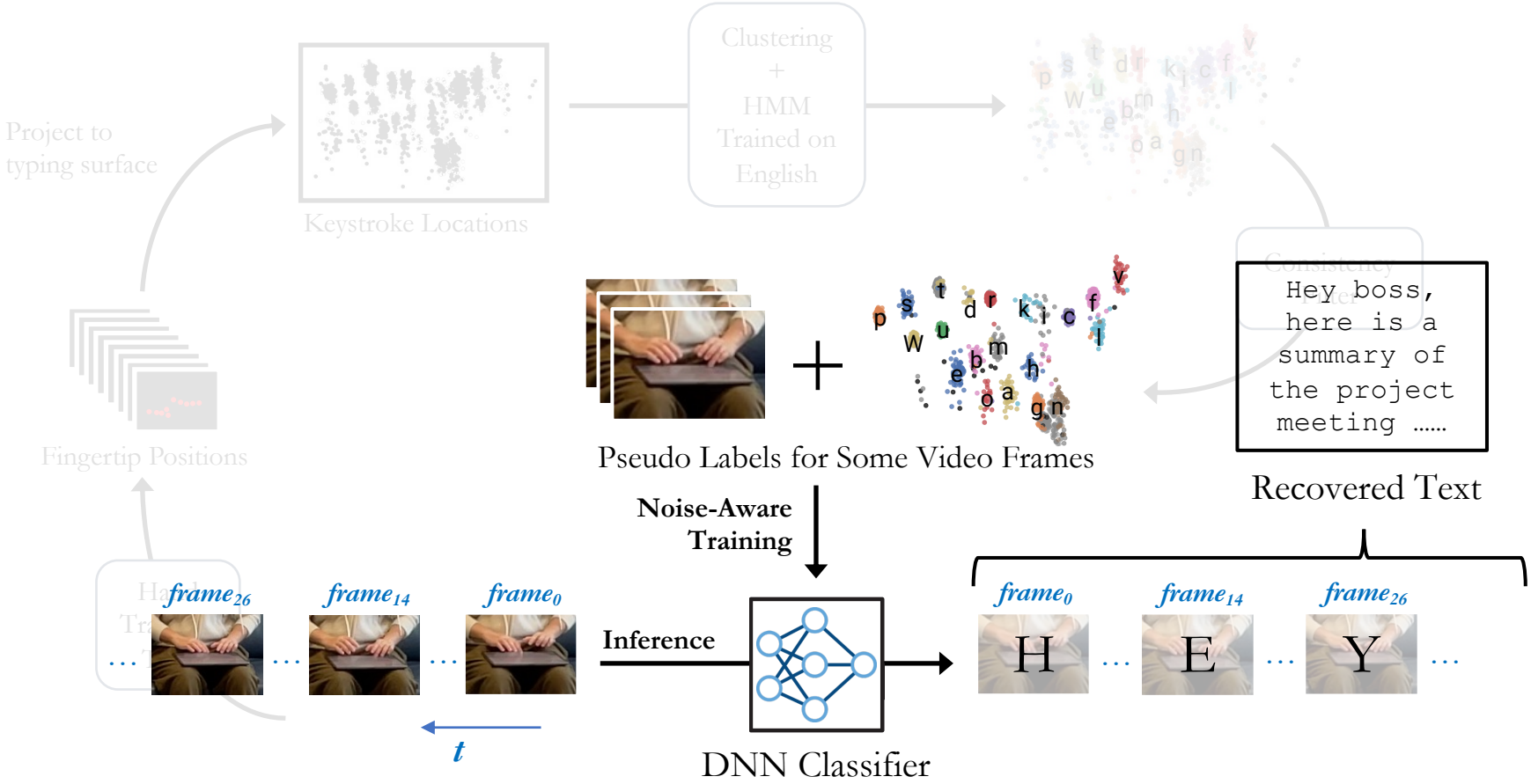


Step 2 – Model Training



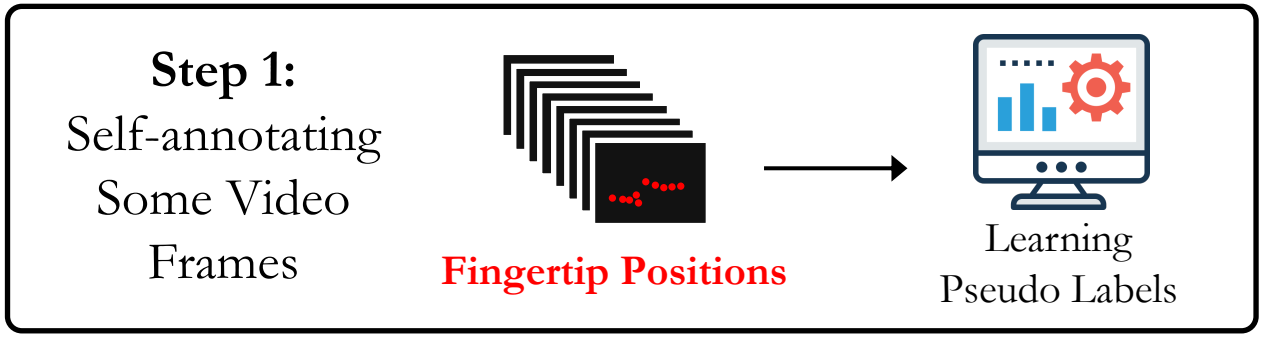
Step 2 – Model Training

See paper for more details



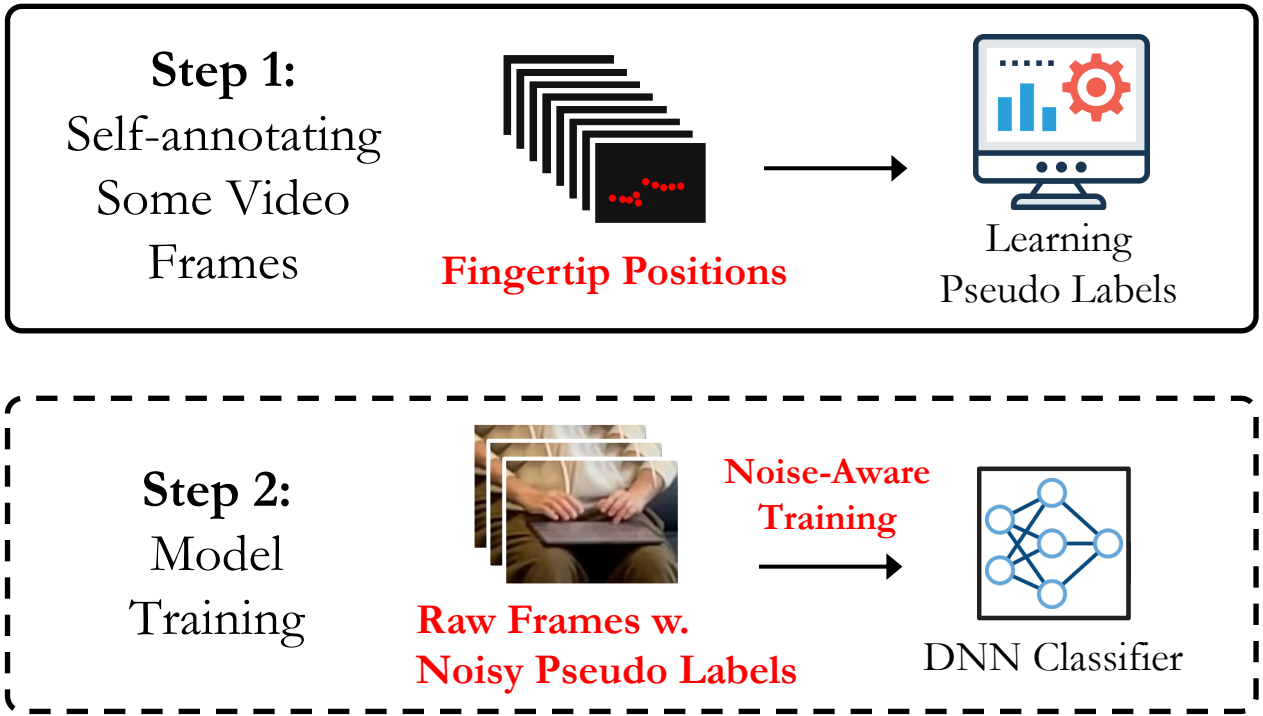
Design Takeaways

Step 1 & 2 use two different data modalities



Design Takeaways

Step 1 & 2 use two different data modalities



Evaluation (Across Users)

- 16 users
- Very different typing styles
- ~500 words email from Enron dataset
  - ~10 mins typing
- iPad keyboard

Evaluation (Across Users)

- 16 users
- Very different typing styles
- ~500 words email from Enron dataset
  - ~10 mins typing
- iPad keyboard

- 3 metrics (original vs. recovered text)
  - CER: character error rate (%) ⬇️
  - WER: word error rate (%) ⬇️
  - Semantic Similarity (%) ⬆️
    - Amount of information recovered

Evaluation (Across Users)

- 16 users
- Very different typing styles
- ~500 words email from Enron dataset
  - ~10 mins typing
- iPad keyboard

- 3 metrics (original vs. recovered text)
  - CER: character error rate (%) ⬇️
  - WER: word error rate (%) ⬇️
  - Semantic Similarity (%) ⬆️
    - Amount of information recovered

User	CER (%) ⬇️	WER (%) ⬇️	Semantic Sim. (%) ⬆️
1	0.7	3.4	99.6
2	1.1	6.0	98.8
3	1.0	3.6	97.4
4	2.3	8.4	97.2
5	3.6	11.2	94.8
6	5.0	12.1	93.5
7	3.9	15.2	91.0
8	5.8	16.5	90.4
9	5.2	14.9	87.6
10	5.1	20.0	84.0
11	8.0	25.5	83.4
12	6.9	17.8	79.9
13	11.6	32.9	71.0
14	12.3	44.8	62.8
15	13.6	35.5	59.0
16	22.8	62.7	14.8

Evaluation (Across Users)

- 16 users
- Very different typing styles
- ~500 words email from Enron dataset
  - ~10 mins typing
- iPad keyboard

- 3 metrics (original vs. recovered text)
  - CER: character error rate (%) ⬇
  - WER: word error rate (%) ⬇
  - Semantic Similarity (%) ⬆
    - Amount of information recovered

User	CER (%) ⬇	WER (%) ⬇	Semantic Sim. (%) ⬆
1	0.7	3.4	99.6
2	1.1	6.0	98.8
3	1.0	3.6	97.4
4	2.3	8.4	97.2
5	3.6	11.2	94.8
6	5.0	12.1	93.5
7	3.9	15.2	91.0
8	5.6	16.5	90.4
9	5.2	14.9	87.6
10	5.1	20.0	84.0
11	8.0	25.5	83.4
12	6.9	17.8	79.9
13	11.6	32.9	71.0
14	12.3	44.8	62.8
15	13.6	35.5	59.0
16	22.8	62.7	14.8

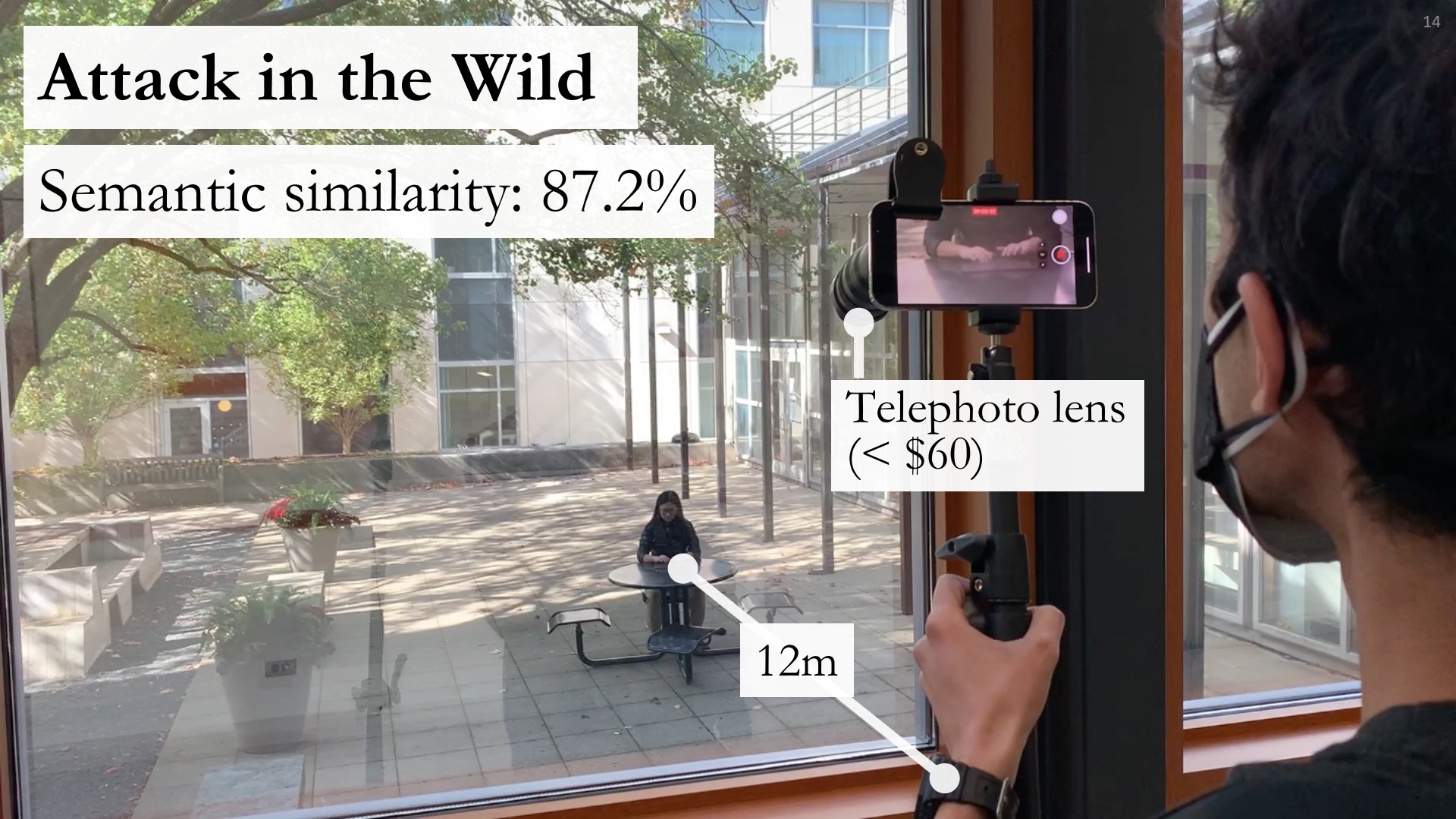
Avg: 6.8%

Avg: 20.7%

Avg: 81.6%

Attack in the Wild

Semantic similarity: 87.2%



No Visual Cue

Equally successful w/ and w/o keyboard

Defense
A physical shield

Project website:

https://sandlab.cs.uchicago.edu/keystroke/

SAND Lab
security, algorithms, networks and data