Preventing fraud with vision in the age of Generative AI

Olivier Koch, VP of Applied AI

AIAI Boston - October 17, 2024

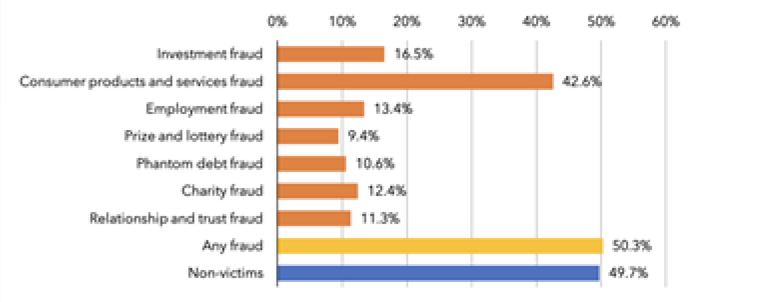
onfido

Online fraud, from individuals to companies & countries

2008

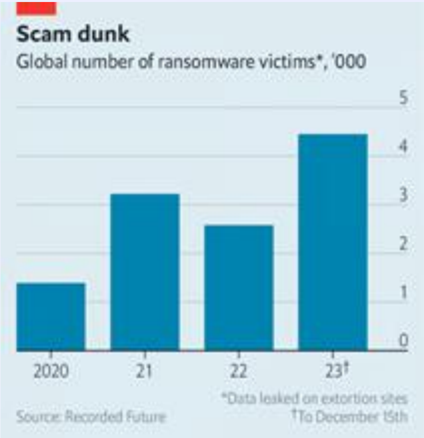
Online fraud, The Economist, Nov'08

2017



Pilot Study to Measure Financial Fraud, Stanford Center on Longevity & FINRA, Feb'17

2023



How ransomware could cripple countries, not just companies, The Economist, Dec'23

"Consumer fraud" costs Americans more than $50bn annually

$1.5M individual ransom payment

Who are we?

Onfido is an online identity verification company.

We let businesses verify the identity of their customers.

Banks

- Revolut
- BARCLAYS
- HSBC
- Sabadell
- bunq
- axiata
- Aspiration
- Millennium bank

Investing

- DRIVEWEALTH
- TRADING 212
- Freetrade
- PensionBee

Lending & Mortgage

- affirm
- zilch
- moneybox
- ZOPA

Payments

- Klarna.
- MANGOPAY
- adyen
- LEMONWAY
- mollie
- AstroPay
- sumup

Gaming

- DRAFT KINGS
- bet365
- William HILL
- MegaBet

Healthcare

- doctor care anywhere.
- babylon
- Teladoc HEALTH
- ZAVA

Travel

- zipcar
- Hertz
- Europcar
- drivy

Other

- DocuSign
- deliveroo
- orange

Onfido’s 3 layers of identity verification

Do you have a **genuine** ID?

1

Are you a **real life** human?

2

Does your face **match** your ID?

3

onfido

Document Verification

+ Thousands of document types
+ Constantly changing attack vectors
+ Variable image quality (API vs SDK)
+ Very low signal-to-noise ratio

onfido

Biometric Verification

- Low friction and accessibility requirements
- Bias reduction
- Deepfakes and injection attacks

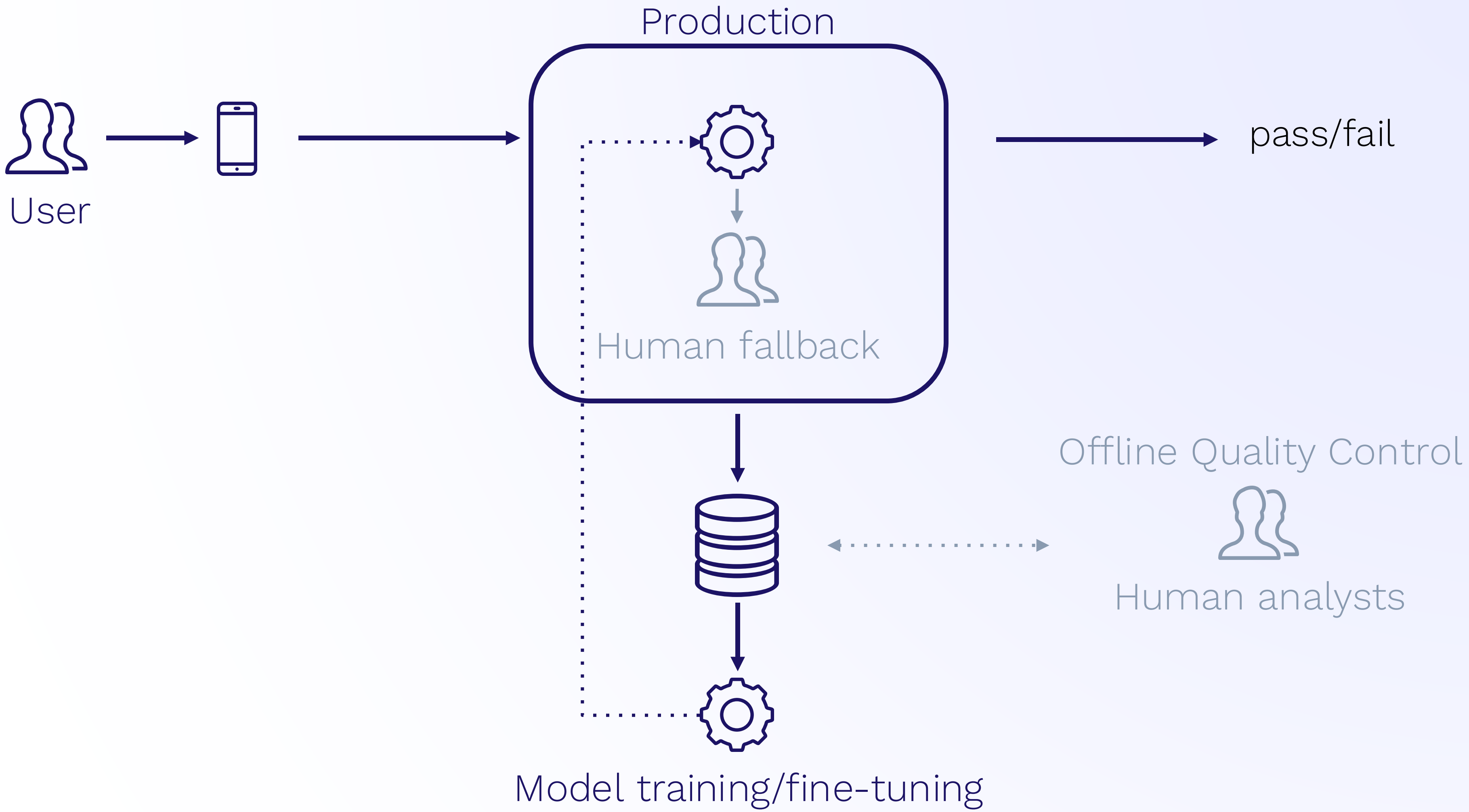
Why online identity verification is hard

Low false alarm

High fraud detection

Global coverage

Constantly evolving attacks



Production
User
pass/fail
Human fallback
Offline Quality Control
Human analysts
Model training/fine-tuning

Automation is key for online identity

Fast

Cheap

Robust

Privacy-friendly

The computer vision pillars of IDV

Face Matching

Extraction

Anomaly Detection

Extraction on thousands of government IDs

Official sample - no PII

Classical extraction methods require human fallback

Template-based
Convnet + LSTM

Hybrid

33.4%

96.4%

Extraction accuracy on 10 fields

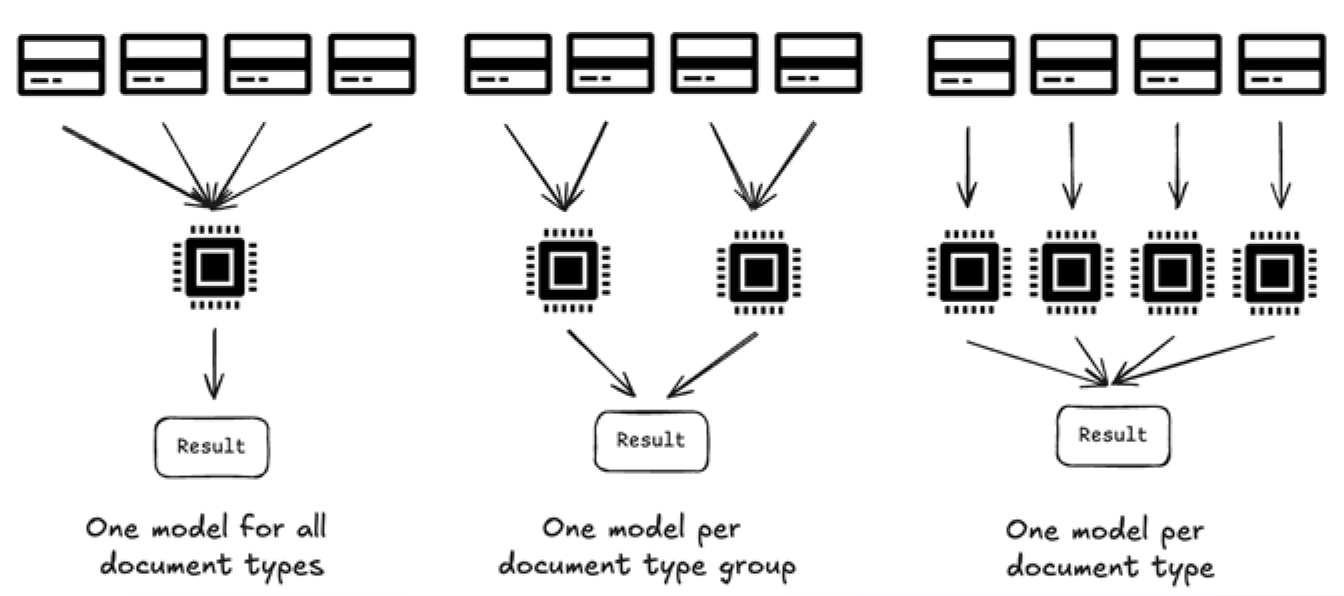
VLMs unlocks much higher extraction accuracy

Template-based Convnet + LSTM 🔗	**VLM-based** (out of the box)	**VLM-based** **optimized**	**Hybrid** (template-based + manual fallback)
33.4%	**76.4%**	**94.6%**	**96.4%**

Extraction accuracy on 10 fields

Leveraging LoRA for cost-efficient extraction

Leveraging large open-source pre-trained models



[Serving models at scale with LoRA](), Martins Bruveris, Oct 2024

The cost effectiveness of in-housing VLMs

In-house

10 g5 GPUs on the cloud -> $75K

3rd-party providers

$0.01 / task -> $1M

for 100M tasks

More control

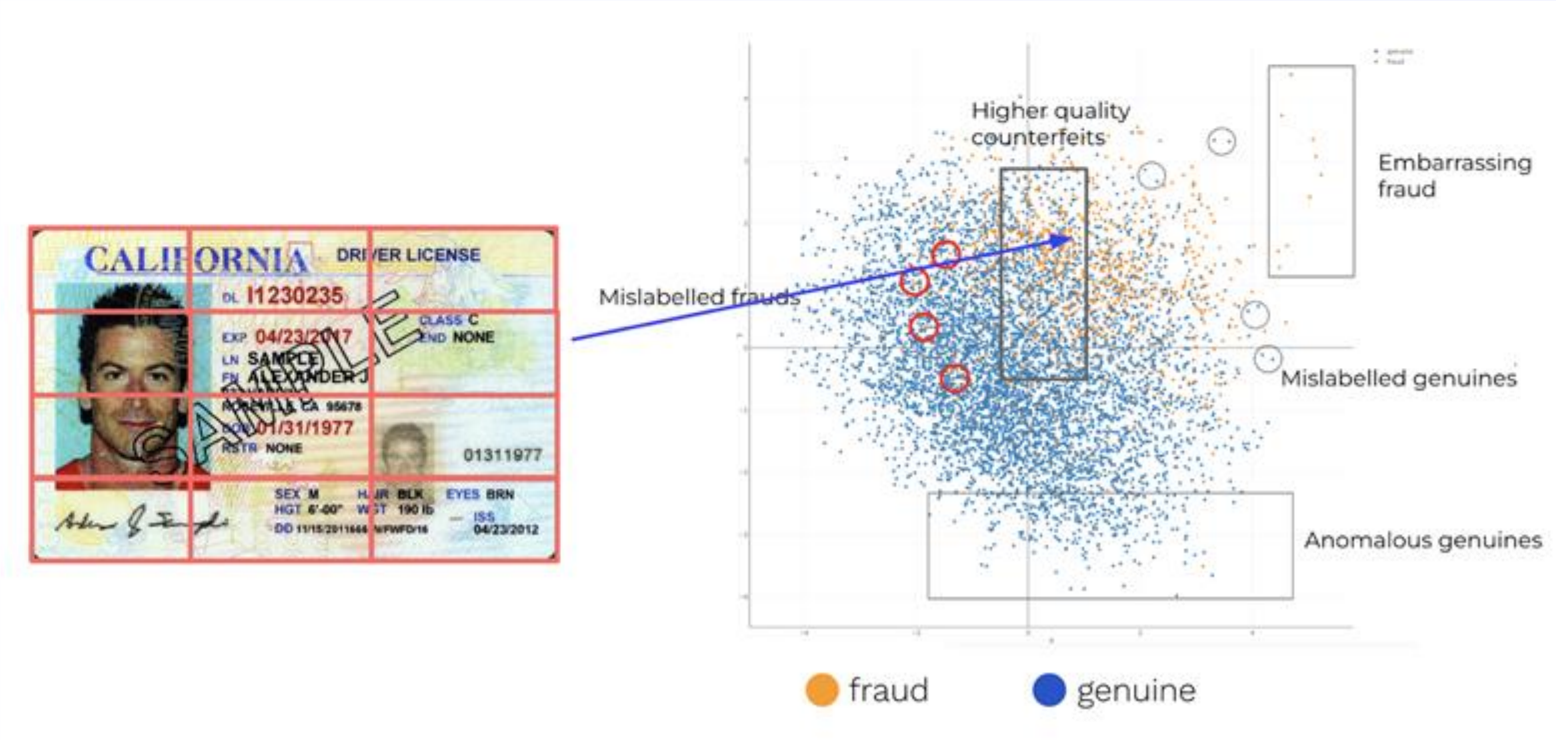
Fine-tune to your need

Increasing regulatory heat

Fraud detection as an anomaly detection problem

- Determine whether a document is fraudulent or not
- Given a large dataset of genuine samples and a *smaller* dataset of frauds
- Across thousands of document types
- And a very large set of anomalies

Vision Transformers for anomaly detection



Source

Vision Transformers for anomaly detection

Leveraging Transformers for visual fraud detection

Left regular auto-encoder. Right a hybrid auto-encoder with a dedicated loss

$$\min_{\theta,\phi} \frac{1}{N_G} \sum_{i=1}^{N_G} ||g_\phi(f_\theta(\mathbf{x}_i)) - \mathbf{x}_i||^2 - \frac{1}{N_F} \sum_{j=1}^{N_F} ||g_\phi(f_\theta(\mathbf{x}_j)) - \mathbf{x}_j||^2$$

Few-shot learning for anomaly detection

Models require hundreds/thousands of samples for training.

Could we make it a few dozens?

Few-shot learning for anomaly detection

Our approach:

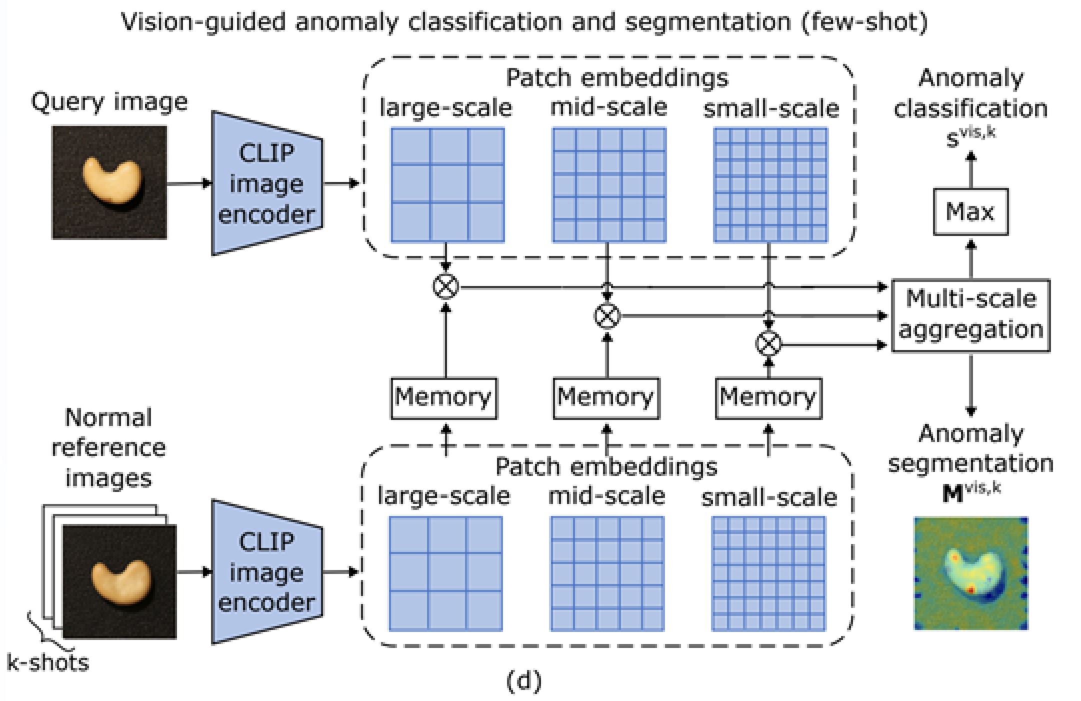
1. Multi-scale GEM embeddings
2. LLM-based prompt ensemble to capture anomaly
3. Zero-shot vision guidance using query image

Outperforms PatchCore and WinCLIP+

On par with AnomalyCLIP, AnomalyGPT and APRIL-GAN w/o auxiliary datasets

FADE Few-shot/zero-shot Anomaly Detection Engine using Large Vision-Language Model, *BMVC 2024, Yuanwei Li, Elizaveta Ivanova, Martins Bruveris*

Few-shot learning for anomaly detection



FADE: Few-shot/zero-shot Anomaly Detection Engine using Large Vision-Language Model, BMVC 2024, Yuanwei Li, Elizaveta Ivanova, Martins Bruveris

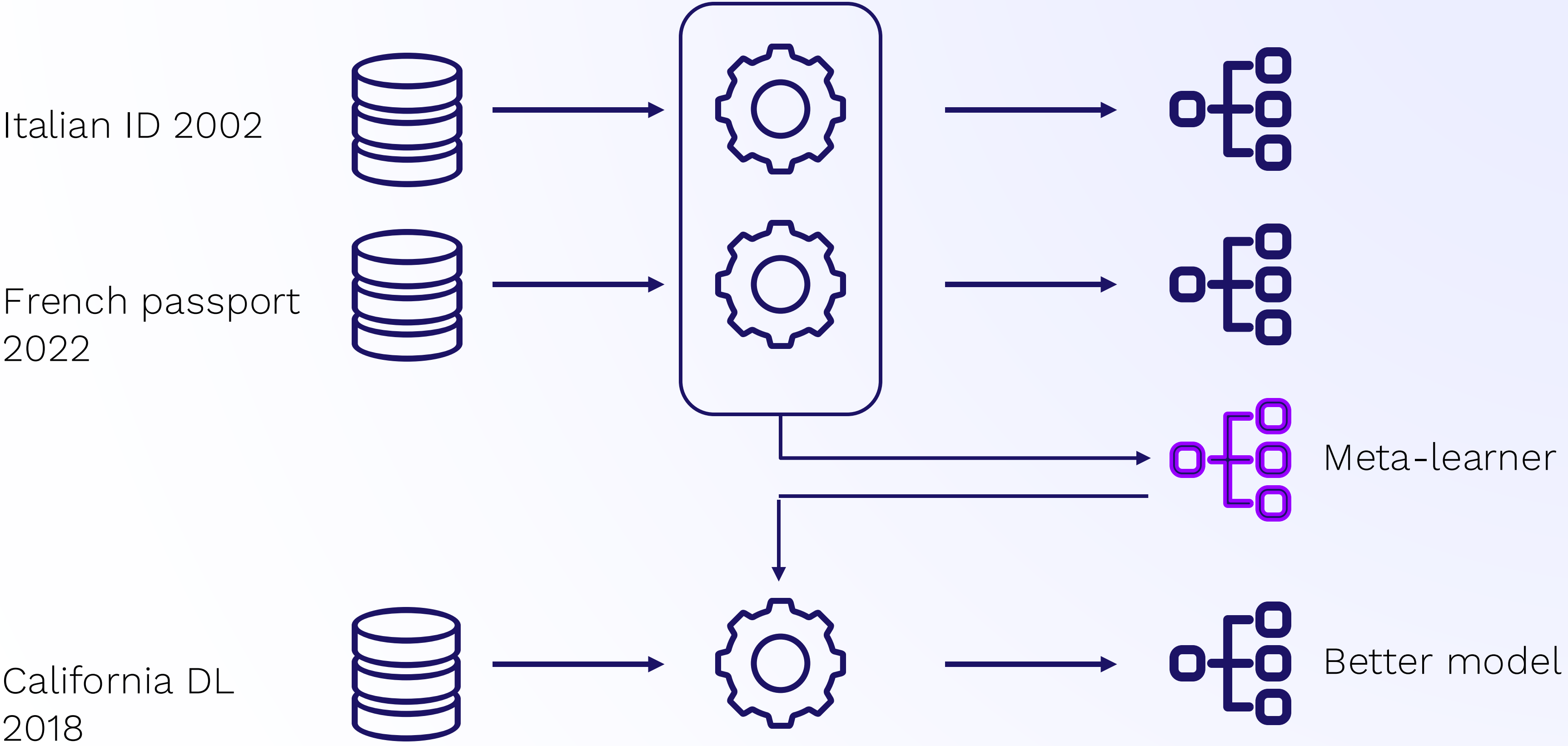
Few-shot learning for anomaly detection

Anomaly Classification		MVTec-AD			VisA		
Setup	Method	AUROC	AUPR	$F1$-max	AUROC	AUPR	$F1$-max
0-shot	WinCLIP	**91.8±0.0**	**96.5±0.0**	**92.9±0.0**	**78.1±0.0**	**81.2±0.0**	**79.0±0.0**
	FADE (ours)	90.0±0.0	95.6±0.0	92.4±0.0	75.6±0.0	78.5±0.0	78.6±0.0
1-shot	PatchCore	83.4±3.0	92.2±1.5	90.5±1.5	79.9±2.9	82.8±2.3	81.7±1.6
	WinCLIP+	93.1±2.0	96.5±0.9	93.7±1.1	83.8±4.0	85.1±4.0	83.1±1.7
	FADE (ours)	**93.9±0.7**	**96.8±0.3**	**94.8±0.2**	**86.7±2.0**	**87.9±1.5**	**84.7±0.8**
2-shot	PatchCore	86.3±3.3	93.8±1.7	92.0±1.5	81.6±4.0	84.8±3.2	82.5±1.8
	WinCLIP+	94.4±1.3	97.0±0.7	94.4±0.8	84.6±2.4	85.8±2.7	83.0±1.4
	FADE (ours)	**95.2±1.0**	**97.6±0.5**	**95.0±0.4**	**89.2±0.4**	**90.2±0.2**	**85.9±0.6**
4-shot	PatchCore	88.8±2.6	94.5±1.5	92.6±1.6	85.3±2.1	87.5±2.1	84.3±1.3
	WinCLIP+	95.2±1.3	97.3±0.6	94.7±0.8	87.3±1.8	88.8±1.8	84.2±1.6
	FADE (ours)	**96.3±0.4**	**98.1±0.2**	**95.5±0.4**	**90.7±0.3**	**91.9±0.4**	**87.0±0.2**

Table 1: Comparison of AC performance on MVTec-AD and VisA. We report the mean and standard deviation over 5 random seeds. Bold indicates the best performance.

FADE Few-shot/zero-shot Anomaly Detection Engine using Large Vision-Language Model, *BMVC 2024, Yuanwei Li, Elizaveta Ivanova, Martins Bruveris*

Meta-learning for low-sample training



[Work presented at ML Prague 2023]

Meta-learning for low-sample training

Zero-shot: MAML outperforms the best pretraining baseline

Few-shot: MAML outperforms significantly in low-data regime, on par in high-data regime

[Work presented at ML Prague 2023]

onfido

Data generation enables faster iteration

Deepfakes are a curse... and a blessing

Synthetic documents

Many problems are still open

- Distillation and transfer learning
- On-device / efficient ML
- Self-supervised learning
- Few-shot learning

We share with the community

FADE: Few-shot/zero-shot Anomaly Detection Engine using Large Vision-Language Model, BMVC 2024, Yuanwei Li, Elizaveta Ivanova, Martins Bruveris

Serving models at scale with LoRA, Martins Bruveris, Oct 2024

Enhancing Deep Learning with Bayesian Inference, Sept'23, Matt Benatan, Jochem Gietema, Marian Schneider

tfimm Star 267

clusterfun Star 14