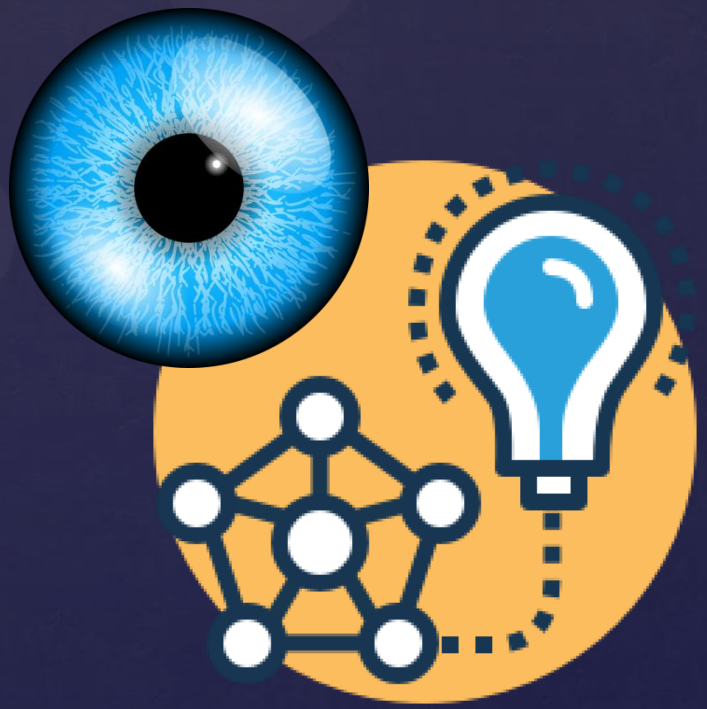
Video Analytics for Security and Compliance

Vishal Kaushal
vkaushal@cse.iitb.ac.in
www.vishalkaushal.in



NCETIS

Advisor: Prof. Ganesh Ramakrishnan

Video is Worth

1.8 million

words per minute

A picture = 1000 words

30,000

words/sec

60

seconds per video

Video shoots at 30 frames/sec

The Flip Side

Time Consuming

Heavy on storage

Video Summarization & Its Flavours

- Watch hours in minutes
- Query Events of Interest
- Generate Useful Statistics

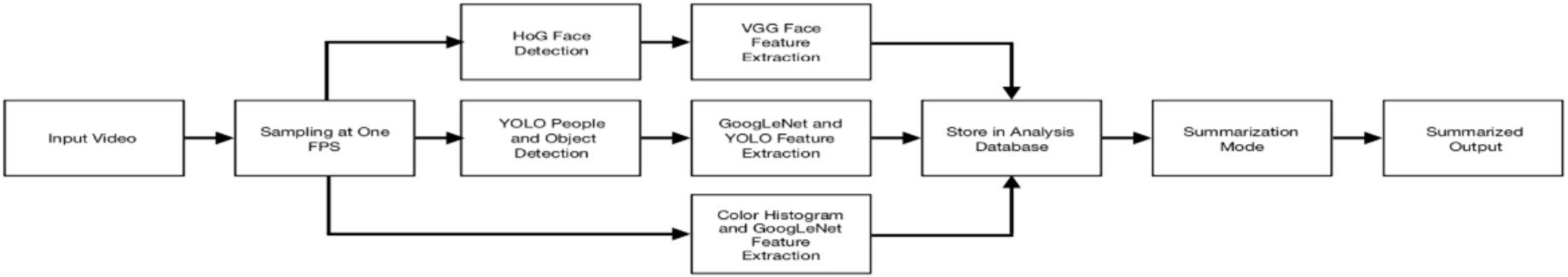
Video Summarization & Its Flavours

- Watch hours in minutes
- Query Events of Interest
- Generate Useful Statistics

Video Summarization & Its Flavours

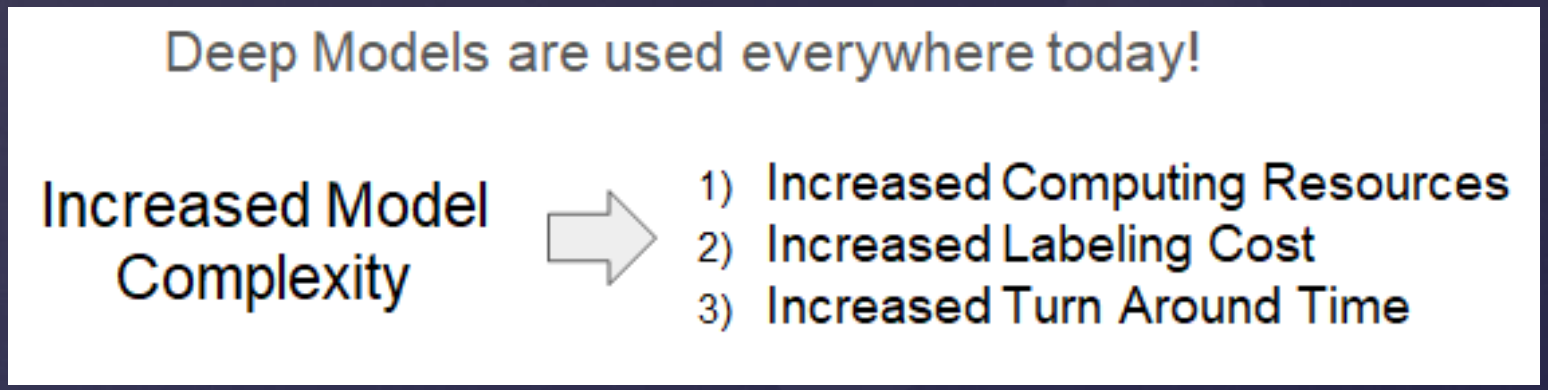
Watch hours in minutes	“Show me a digest of whatever happened today”
Query Events of Interest	“Show me all people wearing yellow shirt” “Alert me as soon as you see a man with covered face”
Generate Useful Statistics	“Number of people in the video” “Motion graph” “Heat map”

A Multi-Faceted Video Summarization System: Tradeoff Between Diversity, Representation, Coverage and Importance (WACV 2019)



- Extractive, Query-based, Entity or Concept based
- In a surveillance video, diversity is more important, while in a movie, coverage and representation becomes more important

Learning From Less Data: A Unified Data Subset Selection and Active Learning Framework for Computer Vision (WACV 2019)



1. Supervised Data Selection for **Quick Training/Inference**
2. Supervised Data Selection for **Quick Hyper-parameter tuning**
3. Unsupervised Data Selection for **Labeling from Video Data**
4. Diversified **Active Learning**

A Framework towards Domain Specific Video Summarization (WACV 2019)

Surveillance Video

Birthday Video

Soccer Video

- **Joint problem** of learning domain specific importance of segments as well as the desired summary characteristic for that domain
- **Ratings** more effective as opposed to binary inclusion/exclusion information
- A **novel evaluation measure**, more naturally suited in assessing the quality of video summary for the task at hand than F1 like measures
- A **gold standard dataset** for furthering research in domain specific video summarization

Summarization via Submodular Functions

For every $X, Y \subseteq \Omega$ with $X \subseteq Y$ and every $x \in \Omega \setminus Y$ we have

$$f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y)$$

Information of Summary Subset S

$$\max_{S \subseteq V, c(S) \leq \mathcal{B}} f(S)$$

Cost of Summary Subset S (e.g. size)

Problem Formulation

Initialization $S \leftarrow \emptyset$.
repeat
Pick an element $v^* \in \text{argmax}_{v \in V \setminus S} \frac{f(v \cup S) - f(S)}{c(v)}$
Update $S \leftarrow S \cup v^*$
until Reaching the budget, i.e., $c(S) > \mathcal{B}$

Greedy Algorithm

Efficient greedy heuristic for optimization [G. Nemhauser, L. Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions, 1978]
Often guaranteed to yield a near optimal solution

Learning Framework

$$o(x_v, y) = w^T f(x_v, y)$$

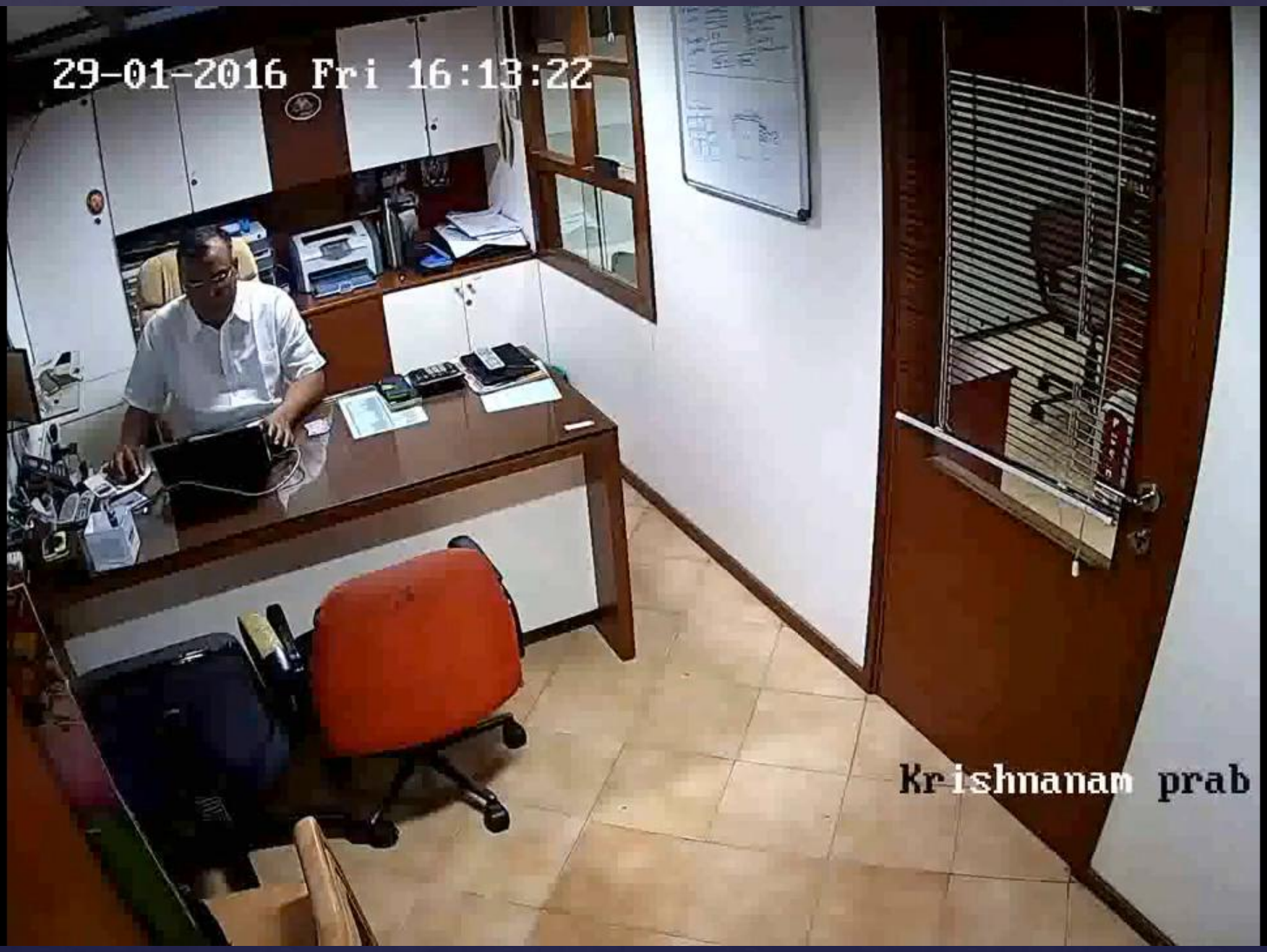
$$y^* = \underset{y \subseteq Y_v, |y| \leq k}{\operatorname{argmax}} \; o(x_v, y)$$

$$S_V(y) = \sum_{x_i \in X_P} |y \cap x_i| * (1 + \frac{|y \cap x_i|}{|x_i|}) * e^{\alpha * rating(x_i)} + \sum_{x_i \in X_R} \min(|y \cap x_i|, \beta) * (1 + \frac{\min(|y \cap x_i|, \beta)}{\min(|x_i|, \beta)}) * e^{\alpha * rating(x_i)} - \sum_{x_i \in X_N} |y \cap x_i| * k$$

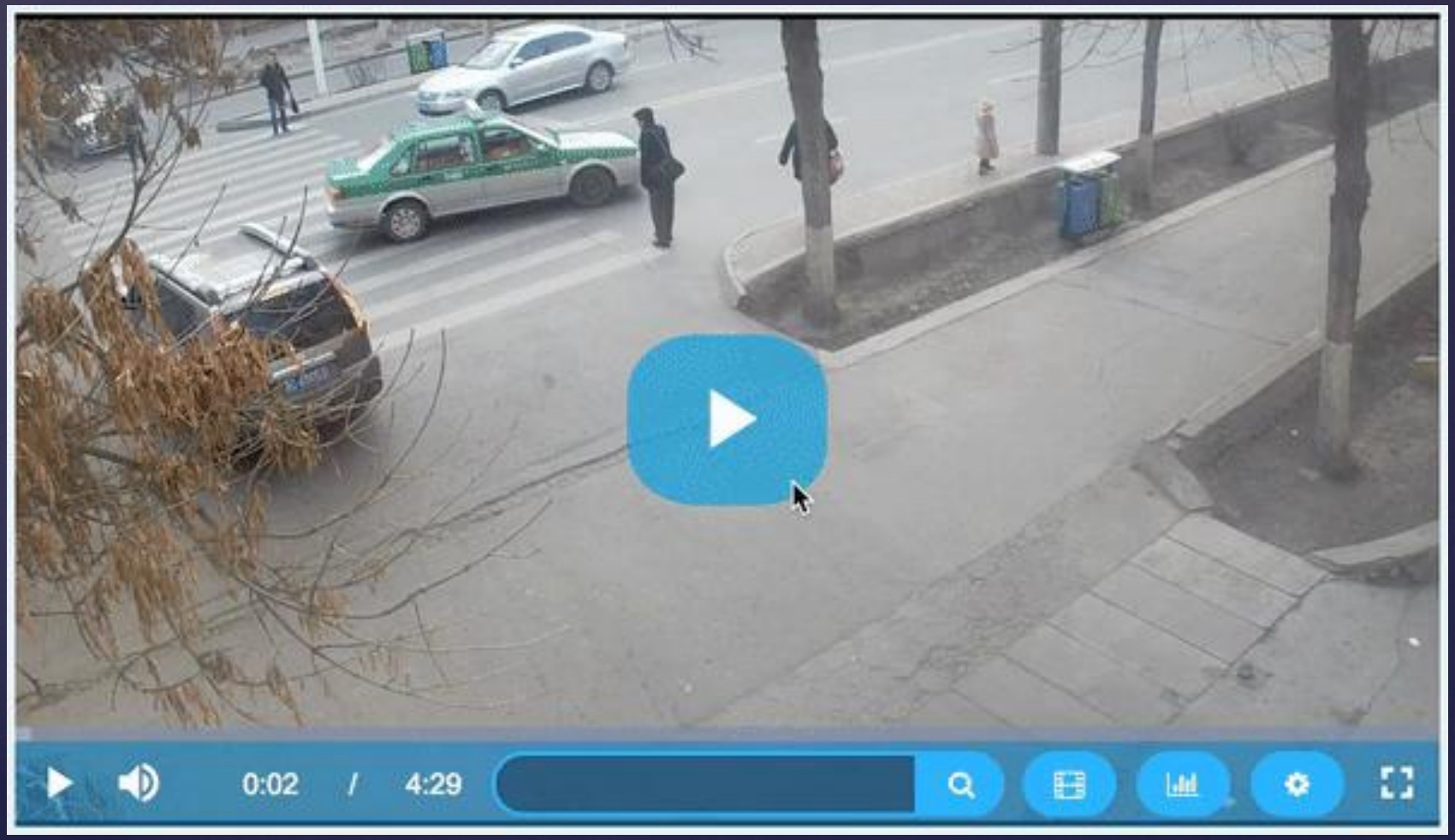
$$\min_{w \geq 0} \frac{1}{N} \sum_{n=1}^{N} L_n(w) + \frac{\lambda_1}{2} ||w_1||^2 + \frac{\lambda_2}{2} ||w_2||^2$$

$$L_n(w) = \max_{y \subseteq Y_v^n} (w^T f(x_v^n, y) + l_n(y)) - w^T f(x_v^n, y_{gt}^n)$$

Watch hours in minutes

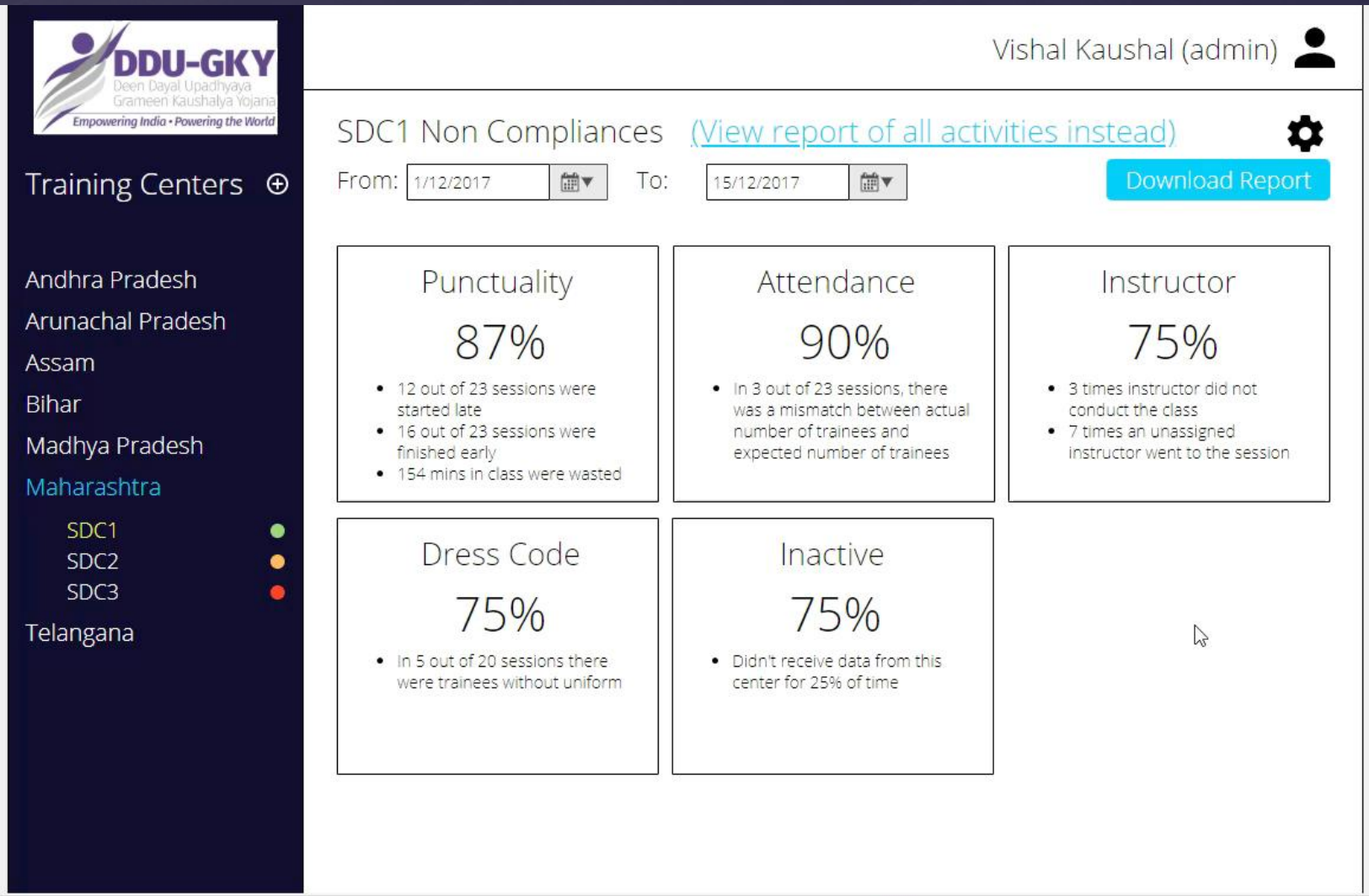


Smart Search



Real-time Alerts

Compliance

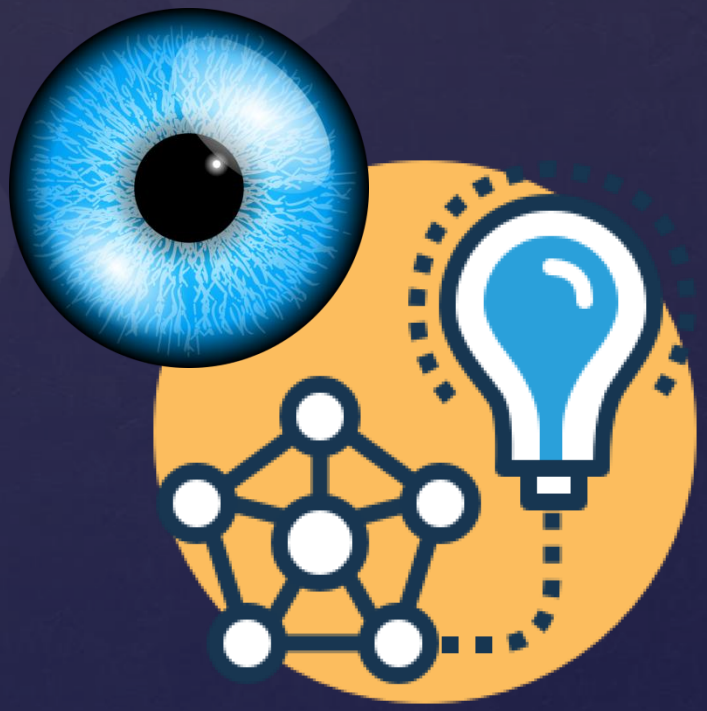

DDU-GKY
Deen Dayal Upadhyaya Grameen Kaushalya Yojana
Empowering India • Powering the World
Training Centers
Andhra Pradesh
Arunachal Pradesh
Assam
Bihar
Madhya Pradesh
Maharashtra
SDC1
SDC2
SDC3
Telangana
Vishal Kaushal (admin)
SDC1 Non Compliances
(View report of all activities instead)
From: 1/12/2017
To: 15/12/2017
Download Report
Punctuality
87%
12 out of 23 sessions were started late
16 out of 23 sessions were finished early
154 mins in class were wasted
Attendance
90%
In 3 out of 23 sessions, there was a mismatch between actual number of trainees and expected number of trainees
Instructor
75%
3 times instructor did not conduct the class
7 times an unassigned instructor went to the session
Dress Code
75%
In 5 out of 20 sessions there were trainees without uniform
Inactive
75%
Didn't receive data from this center for 25% of time

What next?

- Main challenge in video summarization community – benchmarking dataset
- Important practical event of interest for video summarization – masked face
  - Real-time resource constrained masked face detection
- Human interaction and gamification for making video summarization a practical reality
  - Pre-training: For creating labelled data
  - During training: Hyper parameter tuning
  - Post-training: Model customization/localization by human feddback to improve accuracy

Thank You

Vishal Kaushal
vkaushal@cse.iitb.ac.in
www.vishalkaushal.in



NCETIS