IFT
More Information
More Accuracy

ON THE DEVELOPMENT OF A CLASSIFICATION BASED AUTOMATED MOTION IMAGERY INTERPRETABILITY PREDICTION

Hua-mei Chen & Genshe Chen

Intelligent Fusion Technology

Erik Blasch

MOVEJ Analytics

Outline

A. Why imagery interpretability?

B. Approach

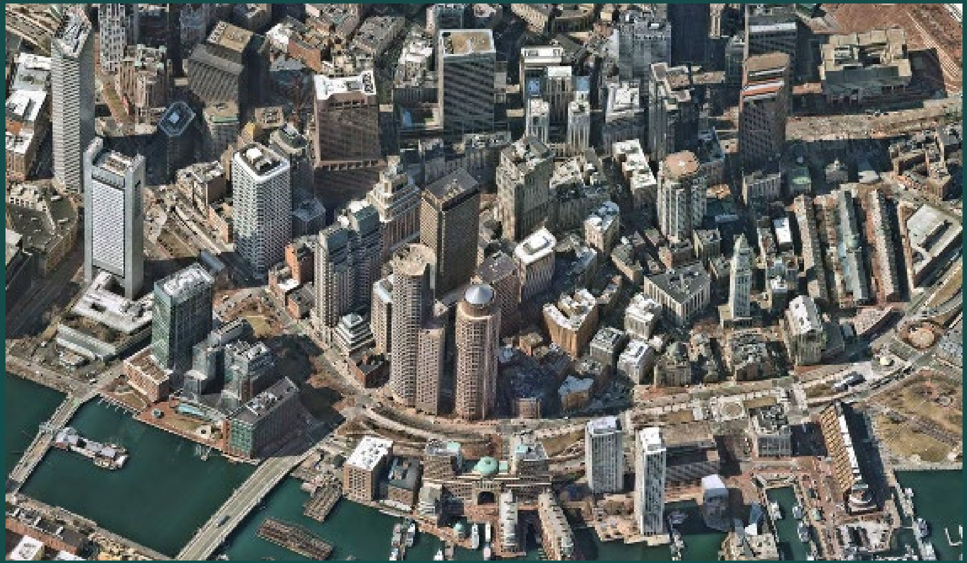
C. Experiments

D. Conclusion

A classification-based motion imagery interpretability prediction estimation accuracy within 0.5 VNIIRS level.

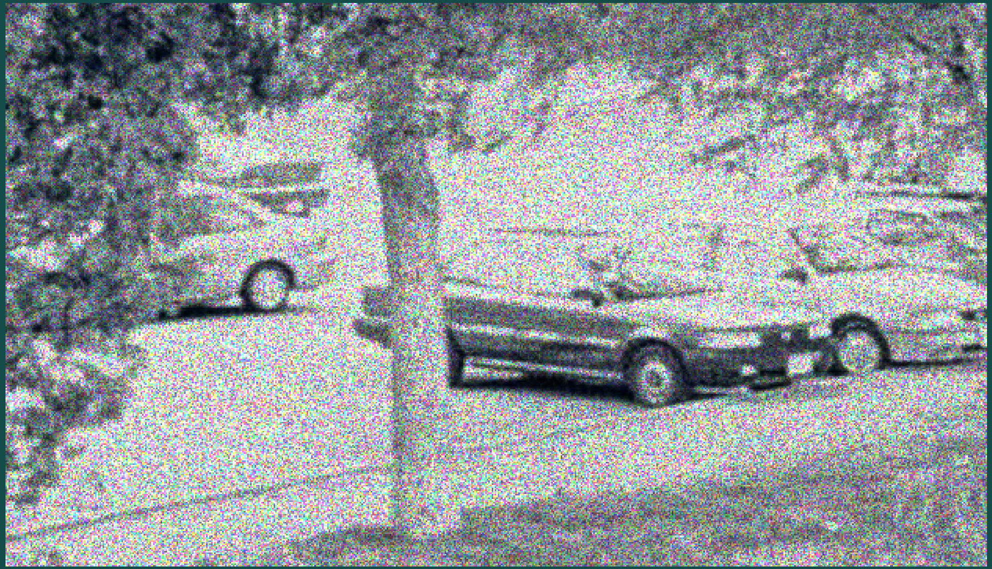
A. Why Imagery Interpretability

- A.1 Imagery quality vs imagery interpretability
  - Quality: overall appearance
  - Interpretability: potential for intelligence task completion



High Quality

Low Interpretability



Low Quality

High Interpretability

A.2 How Do We Quantify Imagery Interpretability?

- Still imagery: National Imagery Interpretability Rating Scale (NIIRS)
- **Motion imagery**: Video-NIIRS or VNIIRS
- Both define a set of different levels of interpretability based on the types of tasks an analyst can perform with imagery of a given rating.
- NIIRS/VNIIRS are subjectively assigned by trained image analysts (IAs)

costly

inefficient

A3. Image/Video Interpretability Estimation Equations

- General Image Quality Equation V4

$$GIQE_4NIIRS = 10.251 - a \cdot log_{10}(GSD_{GM}) + b \cdot log_{10}(RER_{GM}) - 0.656 \cdot H_{GM} - 0.344 \cdot \frac{G}{SNR}$$

- Instantaneous interpretability estimation for the k^th^ frame:

$$\mathbf{I_k = 14 - log_2\,(GSD_k) - log_2(\,1/\,RER_k) - exp(0.5*(PSNR_c\ - PSNR_k))}$$
$$\mathbf{-\,\Delta I_{camera} - \Delta I_{contrast} - \Delta I_{movers} - \Delta I_{artifacts}}$$

Video interpretability equation is ***NOT*** widely accepted as general image quality equations are.

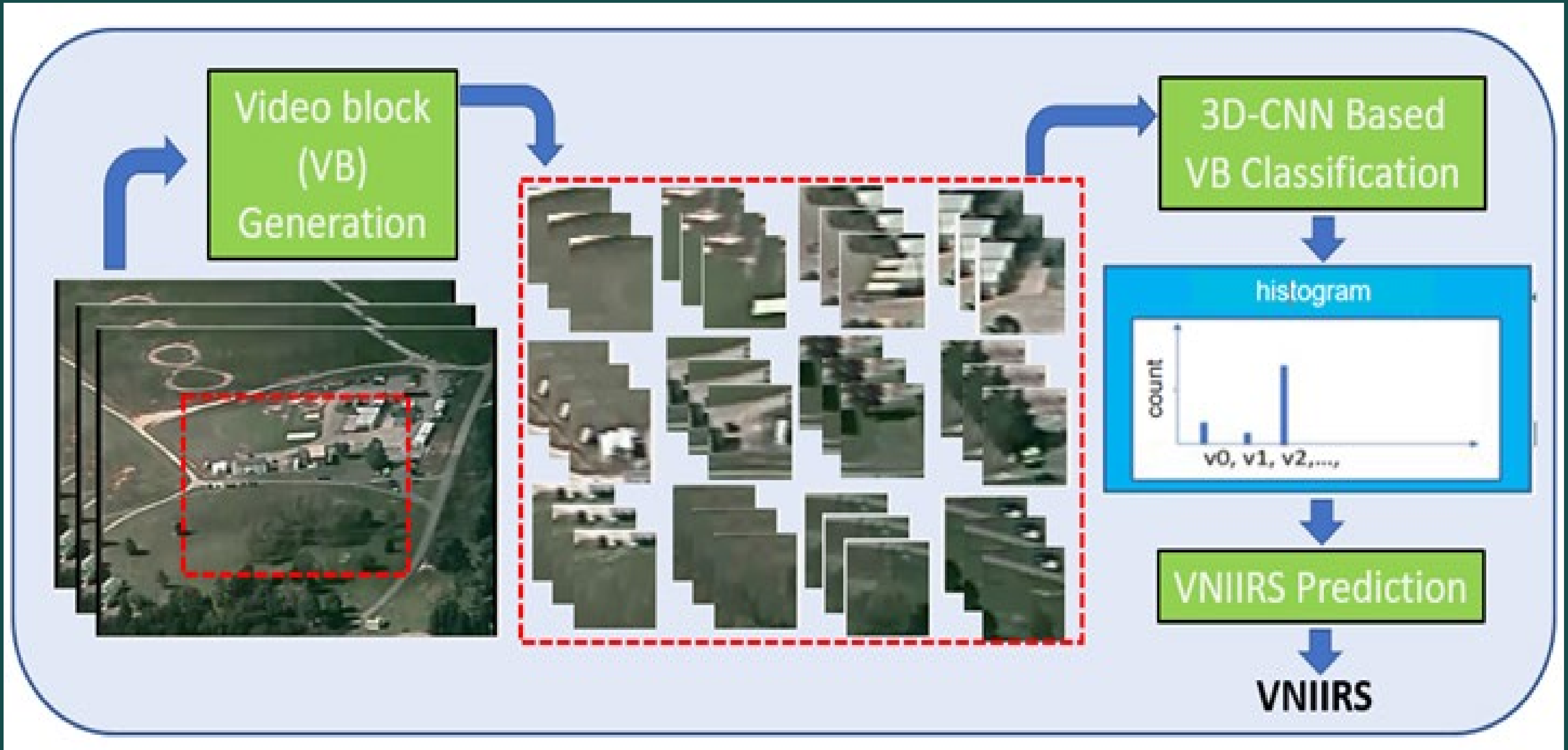
A4. The Need for A Fully Automated Motion Imagery Interpretability Estimation Approach

- Major geospatial intelligence (GEOINT) source.
- Increasing volumes of motion imagery data.
- Lack of trained image analysts.
- Lack of reliable Image Quality Equations (IQE) for VNIIRS estimation

Alternative approaches that do not rely on video interpretability equation are desirable

B. Approach

B.1 Overview



B2. VB Generation

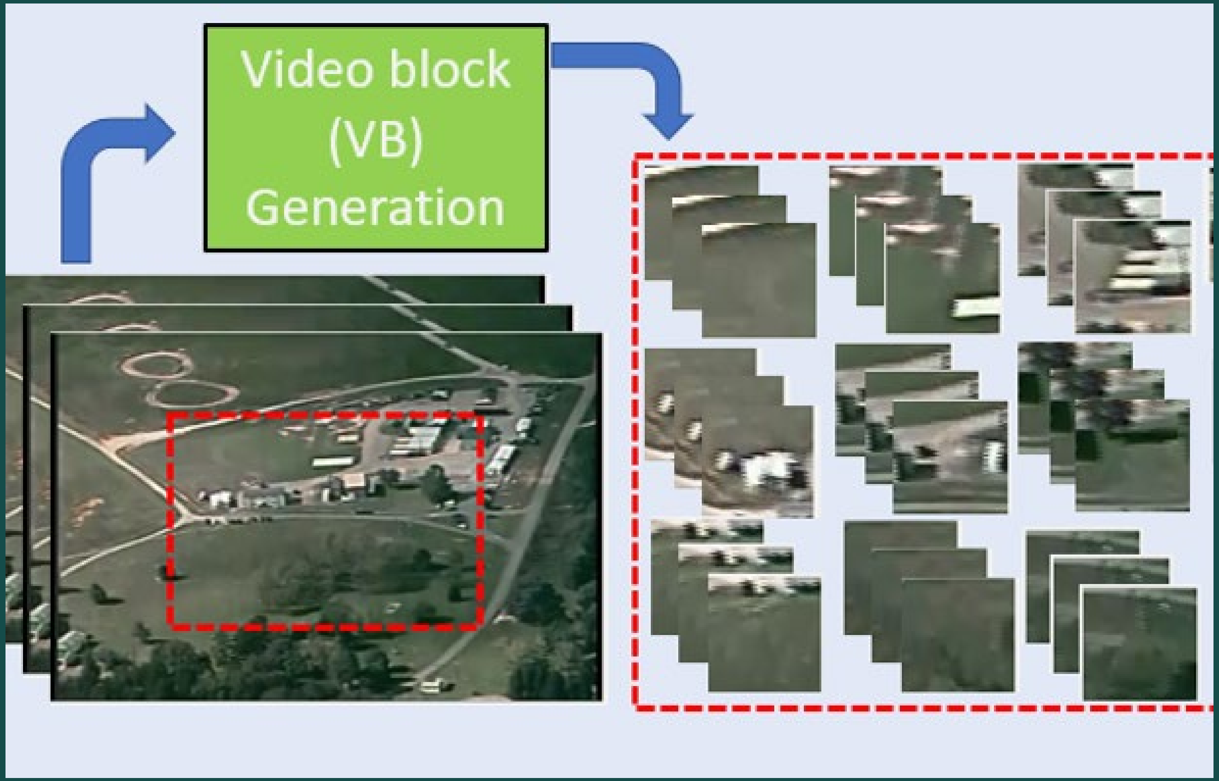
- What is a VB?
- Three tested VB sizes:

  $c \times l \times h \times w = 32\times32\times16\times3,$

  $64\times64\times16\times3,$

  $64\times64 \times 32\times3$

- Input: video clip, output: large number of VBs
- 3d sliding window method is used



B3. VB Selection

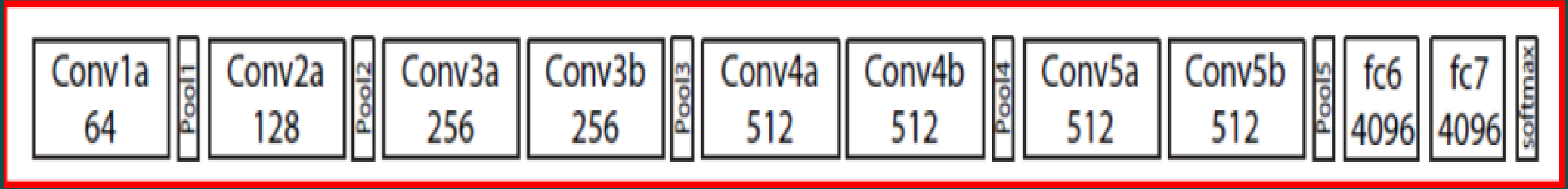
- An informative VB should contain sufficient spatial and temporal variations
- Two VB selection criteria are devised
  - Spatial STD test: $\delta_{spatial} > Th_{spatial}$
  - Temporal STD test: $\delta_{temporal} > Th_{temporal}$
- Our experiment indicated the improvement due to applying the two criteria was not significant.

B4. VB Classification

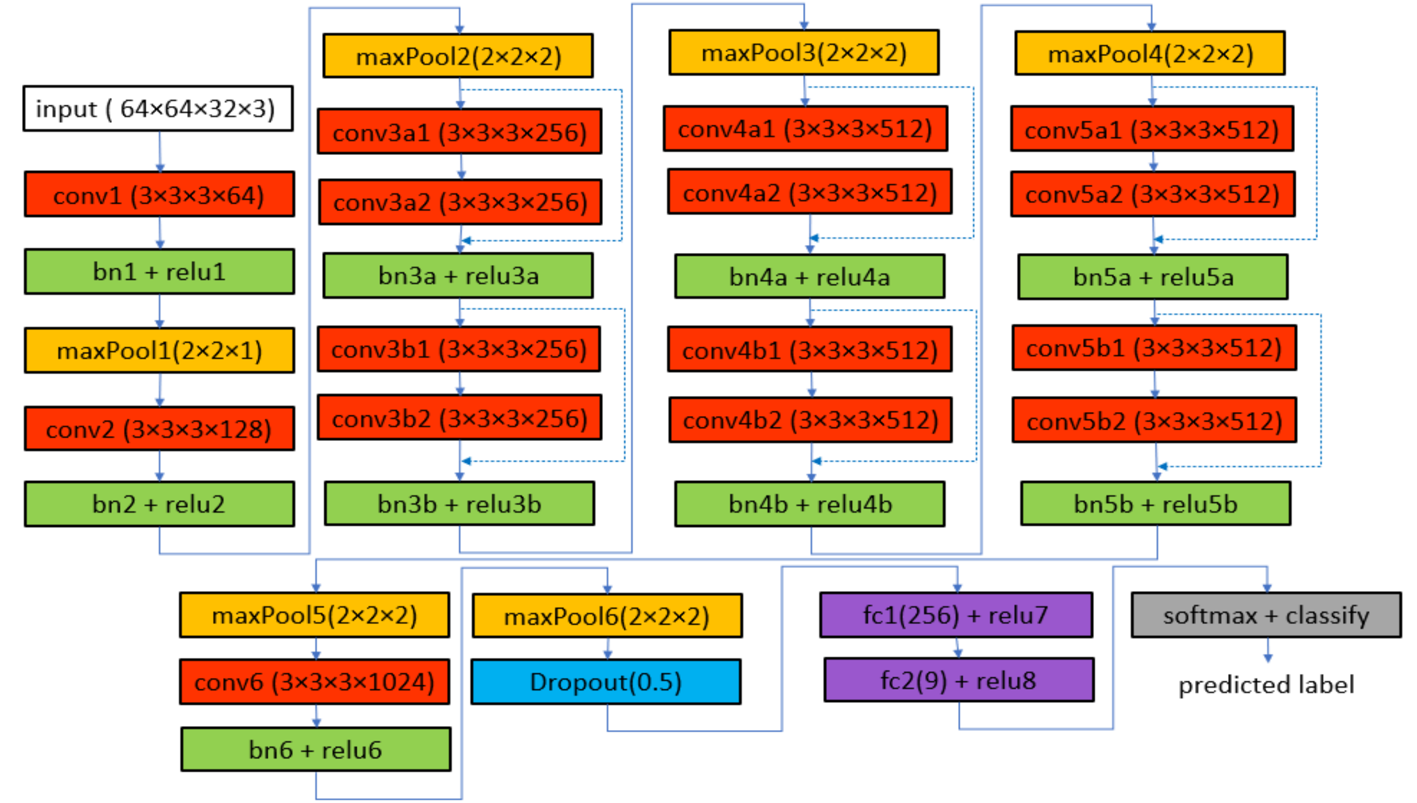
- Based on C3D

input size = 112×112×16×3

- Our implementation
  - Modified the number of convolutional layers to fit specific VB sizes
  - Added batch normalization, dropout, and incorporated residual blocks
  - Class number = 9 , corresponding to VNIIRS 7, 7.5, 8,...,11.



Conv1a 64
Pool1
Conv2a 128
Pool2
Conv3a 256
Conv3b 256
Pool3
Conv4a 512
Conv4b 512
Pool4
Conv5a 512
Conv5b 512
Pool5
fc6 4096
fc7 4096
softmax



input (64×64×32×3)
conv1 (3×3×3×64)
bn1 + relu1
maxPool1(2×2×1)
conv2 (3×3×3×128)
bn2 + relu2

maxPool2(2×2×2)
conv3a1 (3×3×3×256)
conv3a2 (3×3×3×256)
bn3a + relu3a
conv3b1 (3×3×3×256)
conv3b2 (3×3×3×256)
bn3b + relu3b

maxPool3(2×2×2)
conv4a1 (3×3×3×512)
conv4a2 (3×3×3×512)
bn4a + relu4a
conv4b1 (3×3×3×512)
conv4b2 (3×3×3×512)
bn4b + relu4b

maxPool4(2×2×2)
conv5a1 (3×3×3×512)
conv5a2 (3×3×3×512)
bn5a + relu5a
conv5b1 (3×3×3×512)
conv5b2 (3×3×3×512)
bn5b + relu5b

maxPool5(2×2×2)
conv6 (3×3×3×1024)
bn6 + relu6

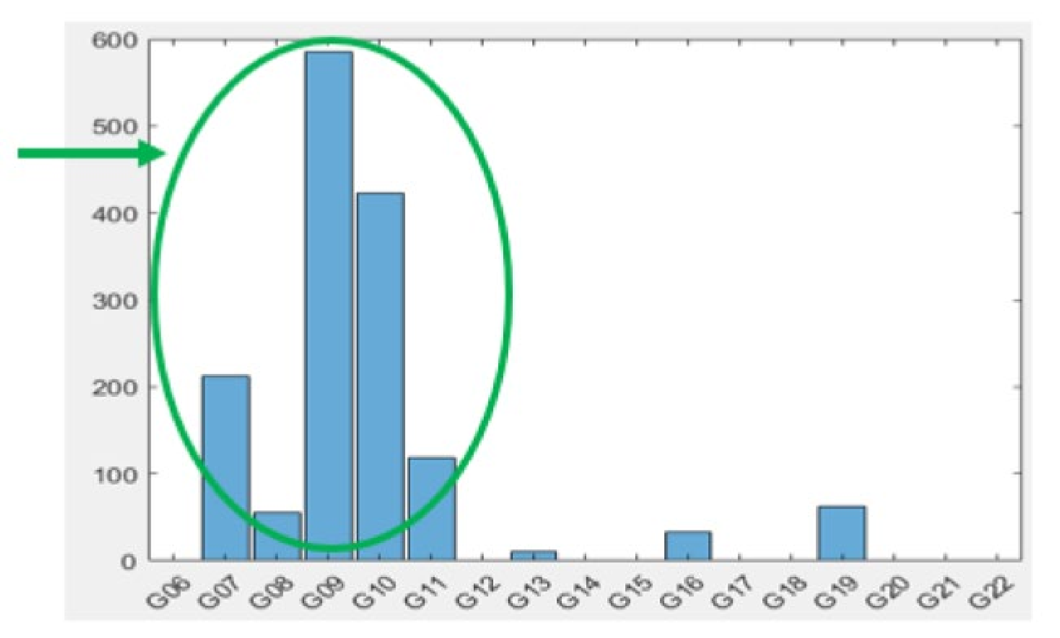
maxPool6(2×2×2)
Dropout(0.5)

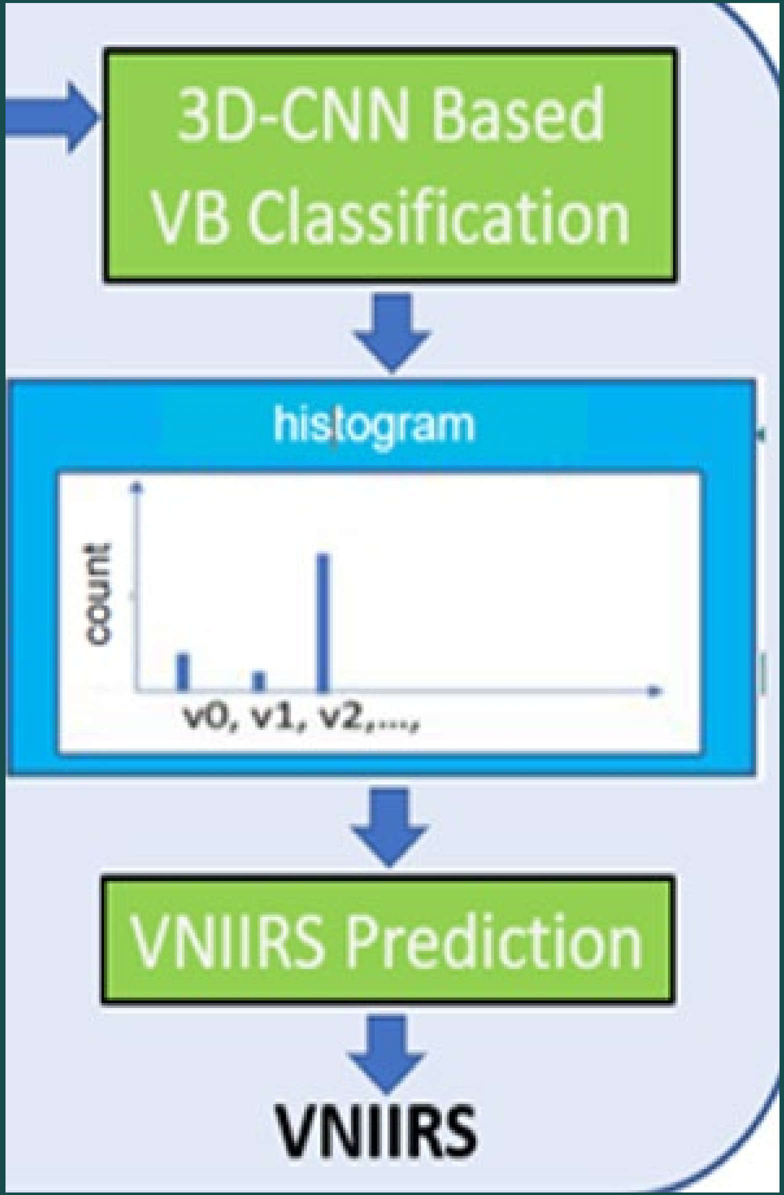
fc1(256) + relu7
fc2(9) + relu8

softmax + classify
predicted label

B5. VNIIRS Prediction

- For each input video clip, the output of VB classification is a **histogram**
- **Estimated VNIIRS = Weighted average**




3D-CNN Based VB Classification

histogram

VNIIRS Prediction

VNIIRS

C. Experiments

C.1 Data Set

- Training set: sixty-six HD aerial video clips
- Test set: ten HD aerial video clips
- VNIIRS range: 7 to 11

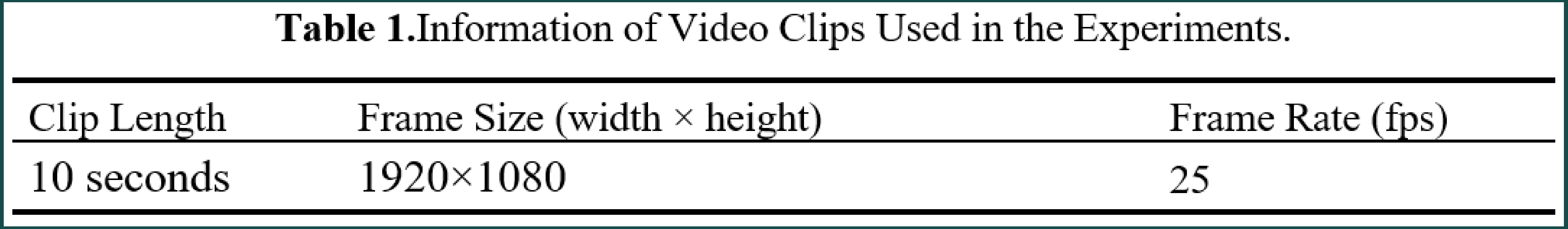


Table 1. Information of Video Clips Used in the Experiments.

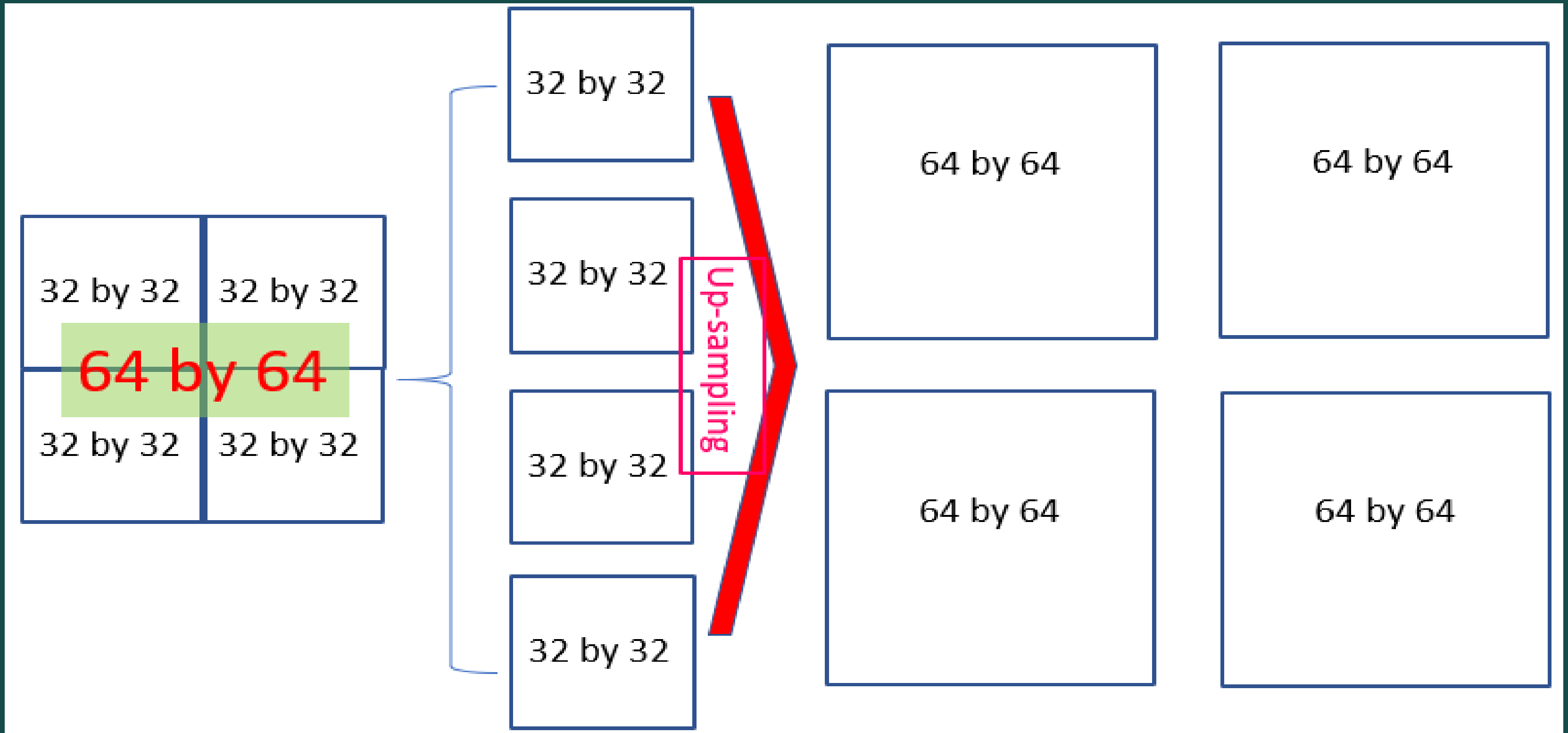
Clip Length	Frame Size (width × height)	Frame Rate (fps)
10 seconds	1920×1080	25

C2. Data Preparation

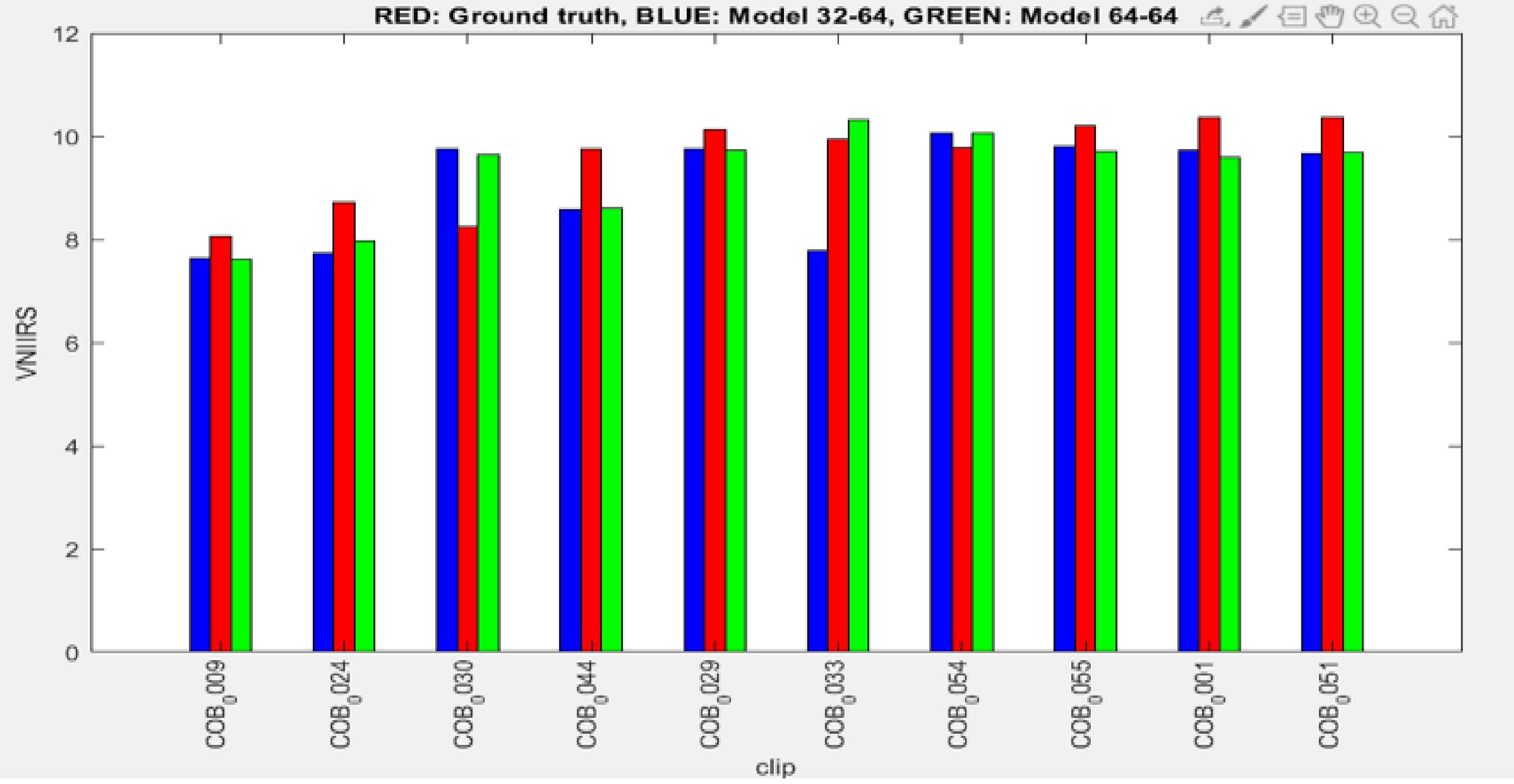
- Each training clip
  - assigned a group label from G14 to G22
  - based on its ground truth VNIIRS level
- Example:
  - VNIIRS = 7.8 ➔ nearest half integer = 8 ➔label G16
  - VNIIRS in G16 ranges from 7.75 to 8.25

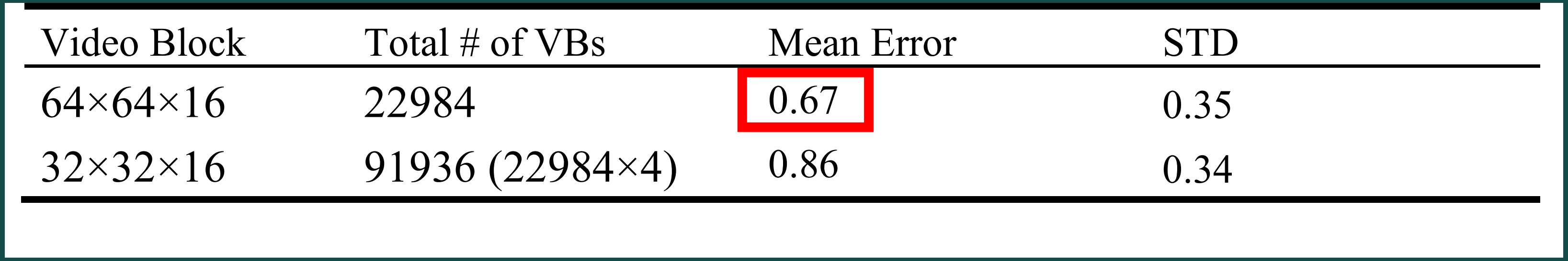
C3. Experiment 1: Performance Comparison of Two Spatial Extents

- Tested two VB sizes: **64**×**64**×16×3 and **32**×**32**×16×3
- Cares are paid to use the same C3D variant and the same training VBs.



C3. Experiment 1: Result

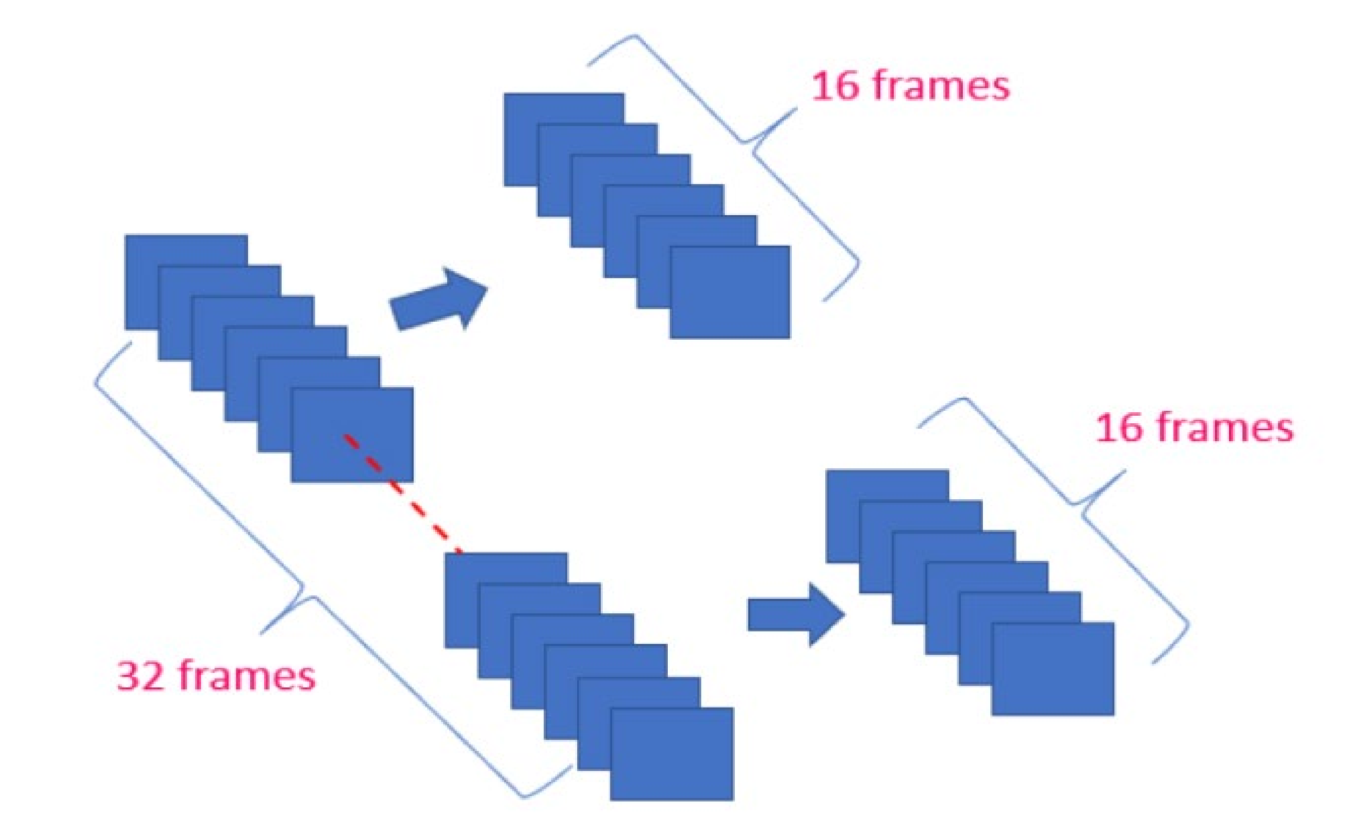




Video Block	Total # of VBs	Mean Error	STD
64×64×16	22984	0.67	0.35
32×32×16	91936 (22984×4)	0.86	0.34

C4. Experiment 2: Performance Comparison of Two Temporal Extents

- Tested two VB sizes: 64×64×**16**×3 and 64×64×**32**×3
- Use the same VBs.
- **Different C3D variants were developed due to different input data sizes**
- Experiment repeated three times (3 classifiers)
- Also tested different number of VB during testing phase.



IFT Proprietary 2020

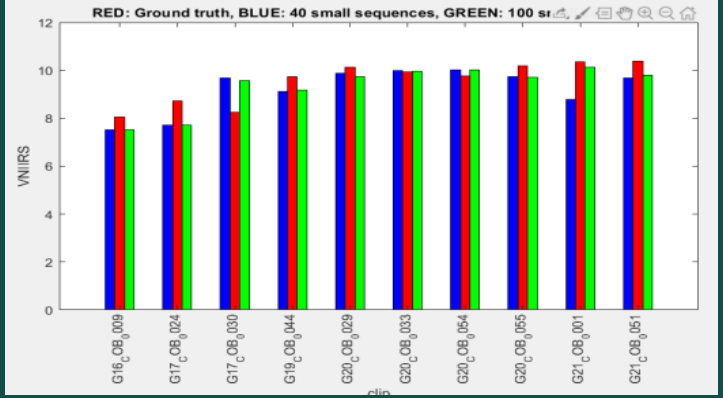
C4. Experiment 2: Graphic Result

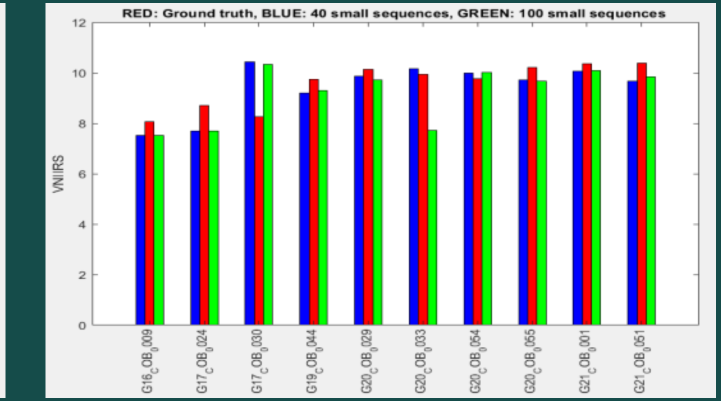
BLUE: 40 test VBs

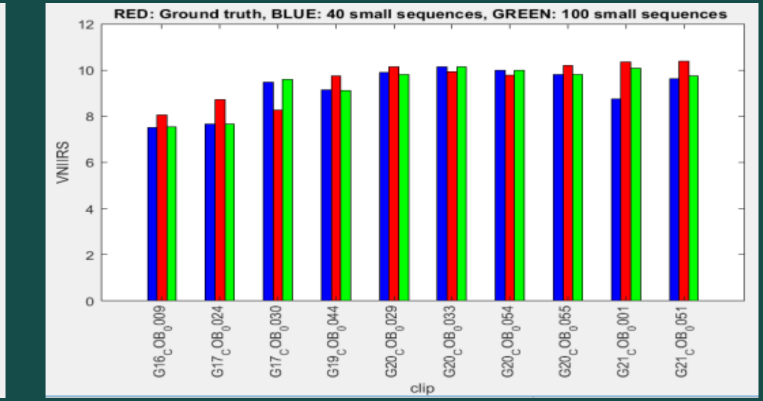
RED: Ground Truth

GREEN: 100 test VBs

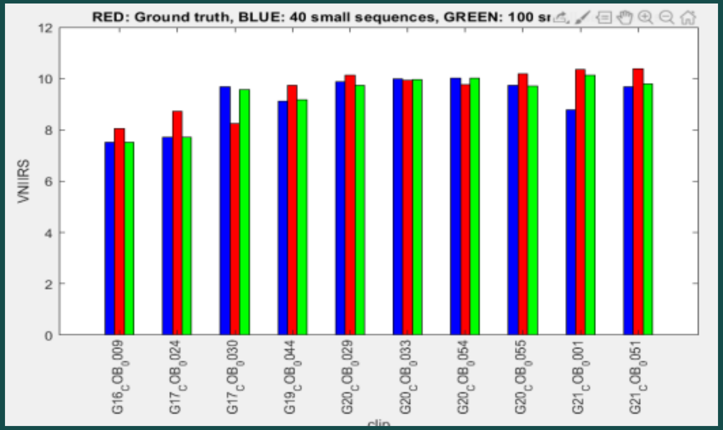
Length 16

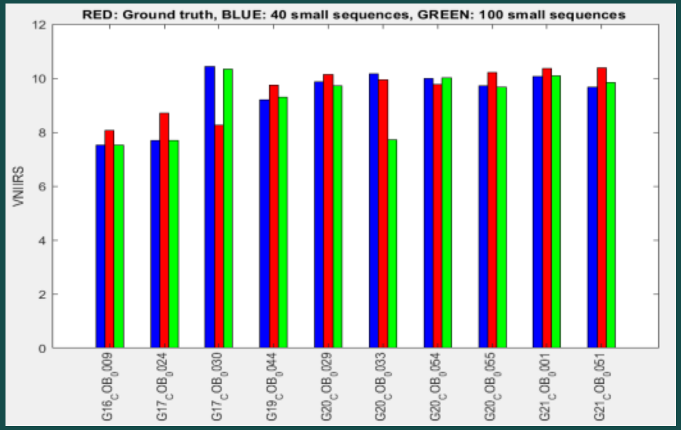

RED: Ground truth, BLUE: 40 small sequences, GREEN: 100 sr

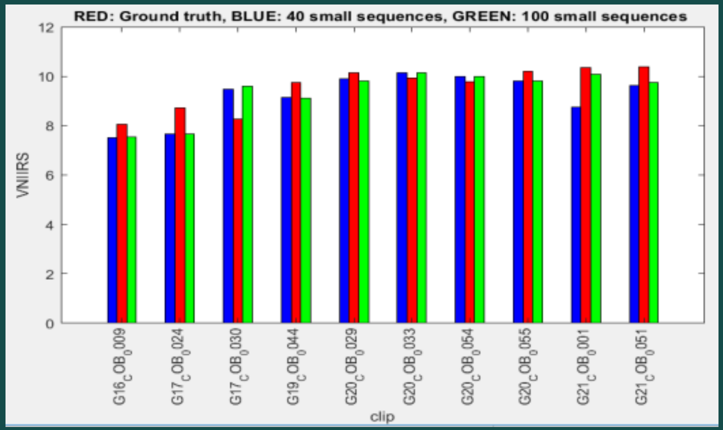

RED: Ground truth, BLUE: 40 small sequences, GREEN: 100 small sequences


RED: Ground truth, BLUE: 40 small sequences, GREEN: 100 small sequences

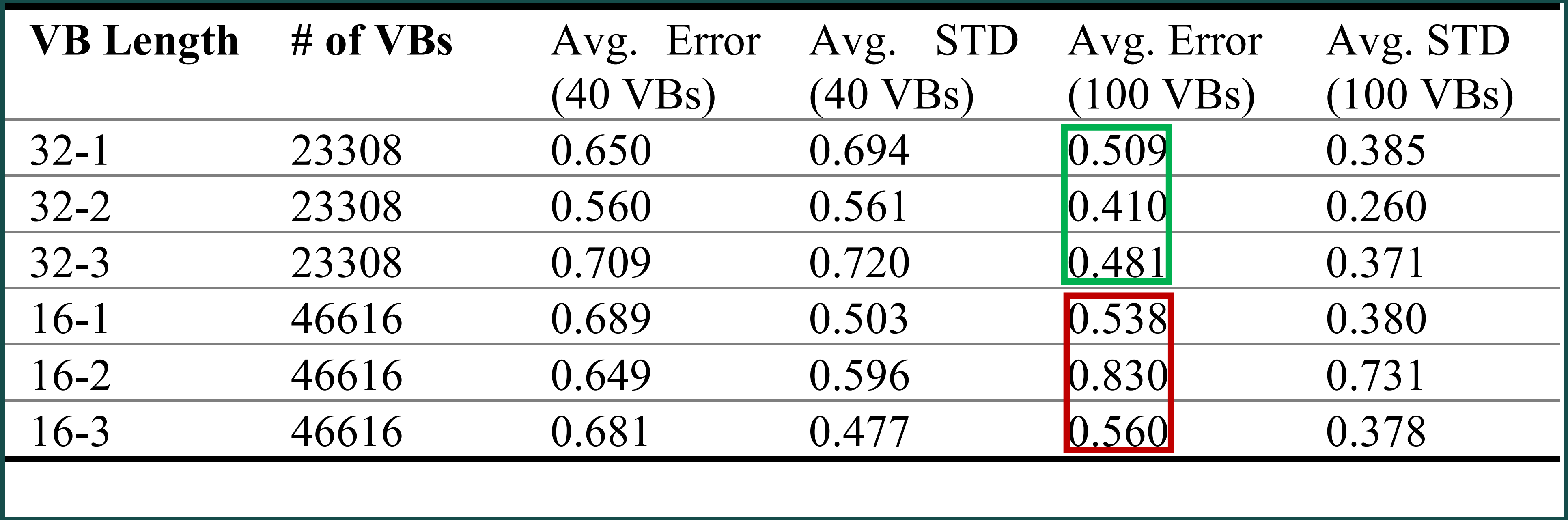
Length 32


RED: Ground truth, BLUE: 40 small sequences, GREEN: 100 sr


RED: Ground truth, BLUE: 40 small sequences, GREEN: 100 small sequences


RED: Ground truth, BLUE: 40 small sequences, GREEN: 100 small sequences

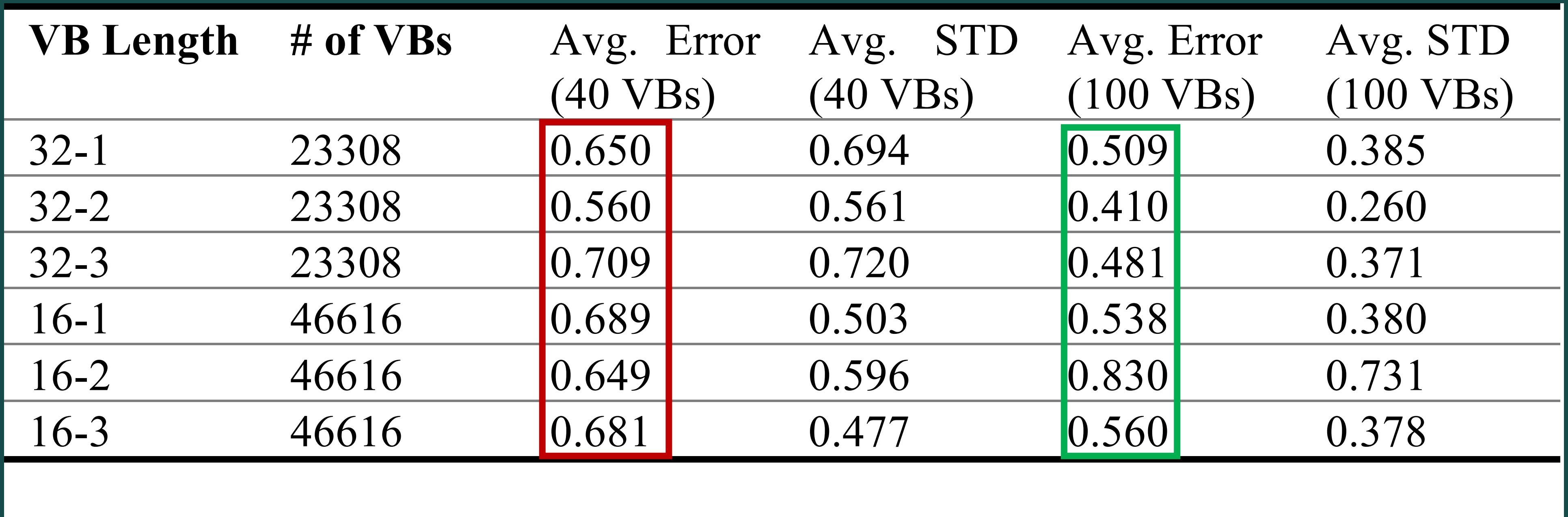
C4. Experiment 2: Numerical Result



VB Length	# of VBs	Avg. Error (40 VBs)	Avg. STD (40 VBs)	Avg. Error (100 VBs)	Avg. STD (100 VBs)
32-1	23308	0.650	0.694	0.509	0.385
32-2	23308	0.560	0.561	0.410	0.260
32-3	23308	0.709	0.720	0.481	0.371
16-1	46616	0.689	0.503	0.538	0.380
16-2	46616	0.649	0.596	0.830	0.731
16-3	46616	0.681	0.477	0.560	0.378

1. Length 32 performs better than length 16
2. Estimation accuracy is about 0.5 VNIIRS level
3. 100 VBs for testing performs 40 VBs for testing in general
4. Performance variation is observed for each C3D variant

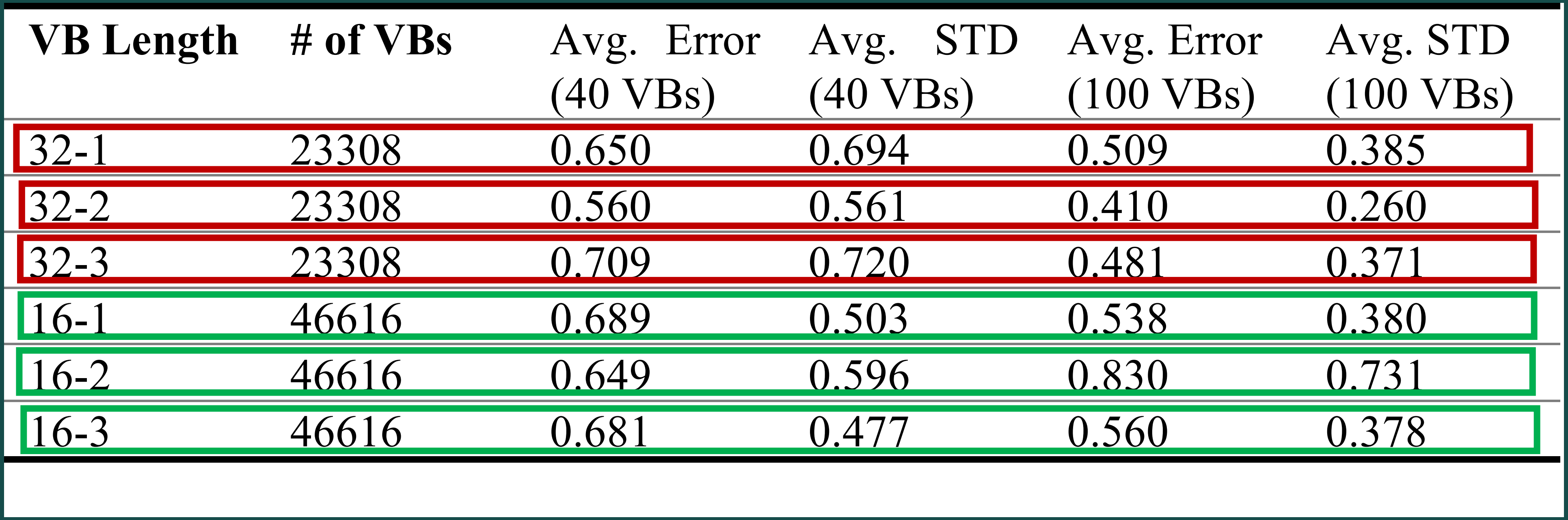
C4. Experiment 2: Numerical Result



VB Length	# of VBs	Avg. Error (40 VBs)	Avg. STD (40 VBs)	Avg. Error (100 VBs)	Avg. STD (100 VBs)
32-1	23308	0.650	0.694	0.509	0.385
32-2	23308	0.560	0.561	0.410	0.260
32-3	23308	0.709	0.720	0.481	0.371
16-1	46616	0.689	0.503	0.538	0.380
16-2	46616	0.649	0.596	0.830	0.731
16-3	46616	0.681	0.477	0.560	0.378

1. Length 32 performs better than length 16
2. Estimation accuracy is about 0.5 VNIIRS level
3. 100 VBs for testing performs 40 VBs for testing in general
4. Performance variation is observed for each C3D variant

C4. Experiment 2: Numerical Result



VB Length	# of VBs	Avg. Error (40 VBs)	Avg. STD (40 VBs)	Avg. Error (100 VBs)	Avg. STD (100 VBs)
32-1	23308	0.650	0.694	0.509	0.385
32-2	23308	0.560	0.561	0.410	0.260
32-3	23308	0.709	0.720	0.481	0.371
16-1	46616	0.689	0.503	0.538	0.380
16-2	46616	0.649	0.596	0.830	0.731
16-3	46616	0.681	0.477	0.560	0.378

1. Length 32 performs better than length 16
2. Estimation accuracy achieved is about 0.5 VNIIRS level
3. 100 VBs for testing performs 40 VBs for testing in general
4. Performance variation is observed for each C3D variant

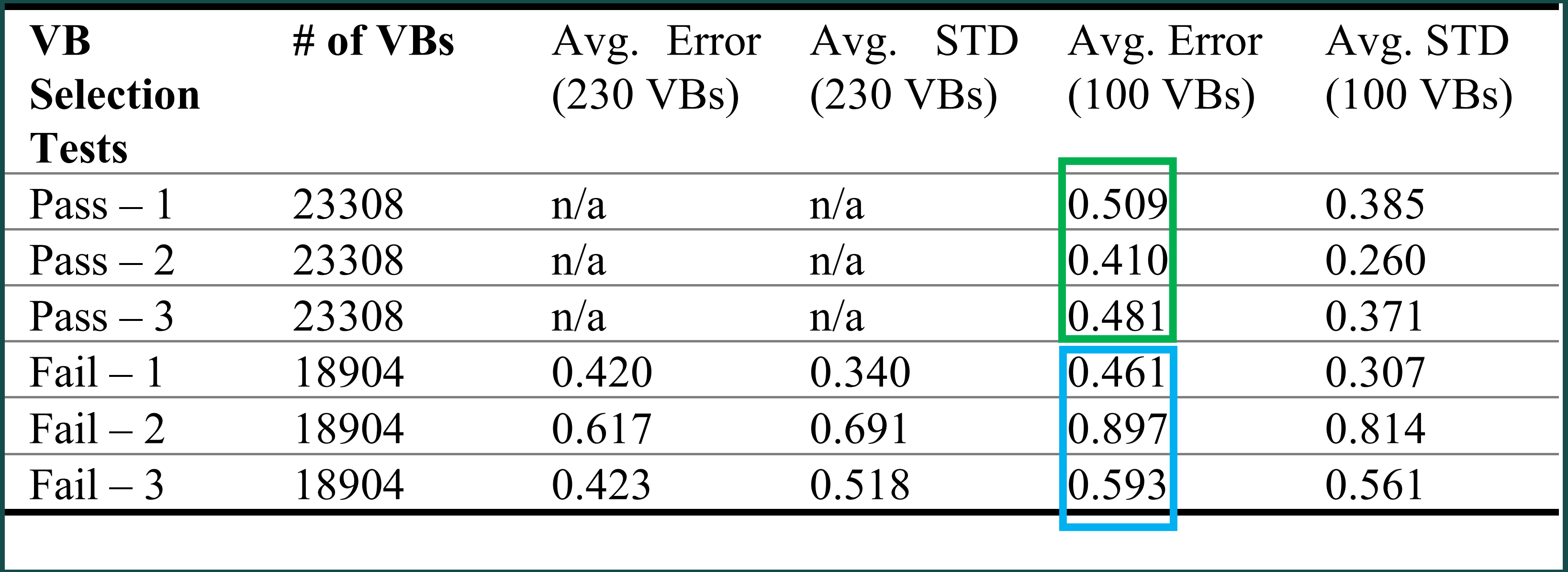
C5. Experiment 3: Test the Effectiveness of Both VB Selection Criteria

- Pass both VB criteria vs Fail both VB criteria

$$\begin{array}{l} \delta_{spatial} > Th_{spatial} \\ \delta_{temporal} > Th_{temporal} \end{array} \quad \text{vs} \quad \begin{array}{l} \delta_{spatial} < Th_{spatial} \\ \delta_{temporal} < Th_{temporal} \end{array}$$

- Tested employing more VBs in the test phase
- VB size was 64×64×32×3

C5. Experiment 3: Numerical Results



VB Selection Tests	# of VBs	Avg. Error (230 VBs)	Avg. STD (230 VBs)	Avg. Error (100 VBs)	Avg. STD (100 VBs)
Pass – 1	23308	n/a	n/a	0.509	0.385
Pass – 2	23308	n/a	n/a	0.410	0.260
Pass – 3	23308	n/a	n/a	0.481	0.371
Fail – 1	18904	0.420	0.340	0.461	0.307
Fail – 2	18904	0.617	0.691	0.897	0.814
Fail – 3	18904	0.423	0.518	0.593	0.561

- No significant difference is observed between using the VBs that pass both criteria and those that fail both criteria
- Using 230 VBs for testing outperformed using 100 VBs for testing

D. Conclusion

- Motion imagery interpretability is about the potential for intelligence task completion
- VNIIRS is defined to quantify motion imagery interpretability
- Subjectively rating motion imagery interpretability is costly and inefficient
- **A classification-based motion imagery interpretability approach is demonstrated**
- **Estimation accuracy within 0.5 VNIIRS level is achieved**
- More data and experiments are needed to consolidate and verify the findings reported in this work

рахмат
danke
謝謝
ngiyabonga
spas
teşekkür ederim
Баярлалаа
mersi
kia ora
barka
welalin
tack
mahalo
tapadh leat
спасибо
faafetai lava
vinaka
blagodaram
dank je
misaotra
paldies
grazzi
matondo
спасибі
thank you
gracias
хвала
kiitos
dankie
nanni
nandri
dhanyavad
maururu
hvala
köszönöm
bedankt
bayarlalaa
gràçie
asante
manana
obrigada
tenki
murakoze
akun
dankon
aĉiŭ
djiere dieuf
tau
mochchakkeram
nkosi
dziękuję
дякую
mamnun
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
chokrane
sobodi
děkuji
sukriya
kop khun krap
obrigado
sagolun
takk
arigatô
dakujem
trugarez
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
grazie
rahmet
terima kasih
kam sah hamnida
rahmat
dhanyavadagalu
shukriya
mercé
мерси
diolch
merci
ευχαριστώ
xièxie
감사합니다
তোমাকে ধন্যবাদ