SkySearch: Satellite Video Search at Scale (Online Appendix)

Minyoung Choe*[1], Geon Lee*[1], Changhun Han*[2], Suji Kim[2], Woong Hu[3], Hyebeen Hwang[3],
Geunseok Park[3], Byeongyeon Kim[4], Hyesook Lee[4], Ha-Myung Park†[2], and Kijung Shin†[1]
[1] *Kim Jaechul Graduate School of AI, KAIST, Seoul, South Korea*, {minyoung.choe, geonlee0325, kijungs}@kaist.ac.kr
[2] *College of Computer Science, Kookmin University, Seoul, South Korea*, {codingnoye, suji2924, hmpark}@kookmin.ac.kr
[3] *Satrec Initiative Co., Ltd., Daejeon, South Korea*, {woonghu, hbh, gspark}@satreci.com
[4] *National Institute of Meteorological Sciences, Jeju, South Korea* {bykim1011, hslee05}@korea.kr

I. Method Details

We present additional details of SkySearch that are not included in the main paper due to space constraints.

Training details on data compression. We detail the training process for the data compression module in SkySearch, which consists of an image encoder and a video encoder. The image encoder captures spatial information using convolutional layers, while the video encoder captures temporal information by processing sequences of image embeddings through a Convolutional RNN. Training is conducted in two stages: first, the image encoder is trained independently, and then the video encoder is trained while keeping the image encoder frozen. Both encoders are optimized using a self-supervised loss function. For a given image (or video) v, we define P_v as the set of positive images (or videos) that are temporally close to v, and N_v as the set of negative images (or videos) that are temporally distant:

$$P_v = \{v' : |t_v - t_{v'}| \leq \Delta\}, \quad N_v = \{v' : |t_v - t_{v'}| > \Delta\}$$

where Δ is a predefined temporal threshold, set to Δ = 8 hours by default. And t_v denotes the timestamp of the image or the initial timestamp of the video. The self-supervised loss is defined as:

$$\mathcal{L}_v = \mathbb{E}_{\substack{p\sim P_v \\ n\sim N_v}} \max(\|f(v) - f(p)\|_2^2 - \|f(v) - f(n)\|_2^2 + \gamma, 0)$$

Here $f(\cdot)$ represents the image or video encoder, and we use γ as 0.5. Equal numbers of positive and negative samples are used during training. The final loss aggregates all images or videos in the database, i.e., $\mathcal{L} = \sum_v \mathcal{L}_v$.

Training details on video prediction. We provide further details about the video prediction module of SkySearch. While SimVP [1] is trained with a pixel-wise MSE loss to predict future video frames, it fails to generate high-quality outputs when applied to high-dimensional satellite videos, as shown in Figures 5 and 6. To address this limitation, SkySearch introduces an adversarial loss by incorporating a discriminator D in addition to the generator (i.e., video predictor) G. Specifically, given an input video v and its ground-truth future video v', a discriminator D is trained to distinguish between the real future video v' and the generated estimated future video $G(v)$. The adversarial loss for the discriminator is defined as:

$$\mathcal{L}_{\text{adv}}^{(D)} = -\mathbb{E}[\log D(v')] - \mathbb{E}[\log(1 - D(G(v))].$$

The generator G is trained to minimize two losses. The first loss is the MSE loss, defined as:

$$\mathcal{L}_{\text{MSE}} = \|v' - G(v)\|_2^2,$$

which aims to make the generated video $G(v)$ closely match the ground-truth future video v' at the pixel level. The second is the adversarial loss, defined as:

$$\mathcal{L}_{\text{adv}}^{(G)} = -\mathbb{E}[\log(D(G(v)))],$$

which encourages the generator to generate video frames that are indistinguishable from real frames by the discriminator.

We further introduce two auxiliary losses to enhance the prediction quality. For the generator, we introduce an orthogonal regularization loss that aims to promote the orthogonality of its parameters, helping to improve generalization and prevent the mode collapse issue. This loss is defined as:

$$\mathcal{L}_{\text{orth}} = \sum_i \|W_i W_i^T - I\|_F^2,$$

where W_i is the weight matrix of the i-th layer in the generator, and I is the identity matrix. For the discriminator, we introduce a feature reconstruction loss to align the latent features of the ground-truth future video v' and the generated video $G(v)$. This loss encourages the discriminator to better align the generated video with the actual video. It is defined as:

$$\mathcal{L}_{\text{feat}} = \|\phi(v') - \phi(G(v))\|_2^2,$$

where $\phi(\cdot)$ is the feature extracted from the discriminator.

To effectively train SkySearch, it is important to balance the contribution of the three main losses. We introduce three hyperparameters, λ_{MSE}, $\lambda_{\text{adv}}^{(G)}$ and $\lambda_{\text{adv}}^{(D)}$, which control the weighting of the MSE loss, the adversarial loss for the generator, and the adversarial loss for the discriminator, respectively. These weights are normalized such that $\lambda_{\text{MSE}} + \lambda_{\text{adv}}^{(G)} + \lambda_{\text{adv}}^{(D)} = 1$. At each training step, we perform biased sampling to select one of the three losses, $\mathcal{L}_{\text{MSE}}$, $\mathcal{L}_{\text{adv}}^{(G)}$, or $\mathcal{L}_{\text{adv}}^{(D)}$ with probabilities

*Co-first authors.
†Co-corresponding authors.

proportional to λ_{MSE}, $\lambda_{\text{adv}}^{(G)}$ and $\lambda_{\text{adv}}^{(D)}$, respectively, to minimize. The two auxiliary loss terms, $\mathcal{L}_{\text{orth}}$ and $\mathcal{L}_{\text{feat}}$, are optimized when the generator or the discriminator are chosen to be trained in the respective training steps. Empirically, we find that $\lambda_{\text{MSE}} = 0.3$, $\lambda_{\text{adv}}^{(G)} = 0.6$ and $\lambda_{\text{adv}}^{(D)} = 0.1$ provide a good balance for effective training.

II. Additional Empirical Results

We present additional experimental results of SkySearch.

Extra similar video search results. As shown in the numerical results in the main paper, SkySearch consistently outperforms ResNet and EfficientNet (both pre-trained and fine-tuned) in both short-term and long-term accuracy. Notably, SkySearch with video prediction achieves the highest accuracy for long-term retrieval, while SkySearch without video prediction excels in short-term accuracy. To complement these numerical results with qualitative evidence, we present additional long-term search results comparing SkySearch to the fine-tuned EfficientNet, the best-performing baseline. To provide a more detailed view of the long-term (24-hour) retrieval, four representative frames from each retrieved 24-hour video are included. As illustrated in Figures 1 and 2, SkySearch with video prediction retrieves videos that are more similar to the query and its future developments compared to both the baseline and SkySearch without video prediction. These results demonstrate the unique ability of SkySearch with video prediction to retrieve videos that not only align with the query but also capture its anticipated future evolution.

Extra results with alternative ranking methods. In SkySearch, once the set C of similar video candidates is retrieved, they are ranked based on their embedding distances to the query embedding in the latent space by default. For enhanced precision, the ranking can be refined using more computationally intensive similarity or distance measures, such as LPIPS [2], FSIM [3], and SSIM [2]. To illustrate this refinement, we provide additional search results using LPIPS, FSIM, and SSIM to select the best match among 50 similar video candidates retrieved by SkySearch with the video prediction module. As shown in Figures 3 and 4, all the retrieved videos are visually similar to the query video. However, results can be further refined using LPIPS, FSIM, and SSIM for improved precision. It is important to note that while embedding distances are computationally efficient, they may result in less precise retrievals compared to these more advanced measures. On the other hand, refining results with LPIPS, FSIM, and SSIM requires significant computational resources, including loading high-resolution images and performing detailed similarity calculations.

Extra video prediction results. In SkySearch, video prediction is used for query video augmentation and visualization. In Figures 5 and 6, we present additional prediction results comparing the video prediction module in SkySearch with the baseline SimVP [1]. SkySearch's video prediction module, enhanced with adversarial learning, yields more accurate and sharper video predictions.

References

[1] Z. Gao, C. Tan, L. Wu, and S. Z. Li, "Simvp: Simpler yet better video prediction," in CVPR, 2022.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," TIP, vol. 13, no. 4, pp. 600–612, 2004.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," TIP, vol. 20, no. 8, pp. 2378–2386, 2011.

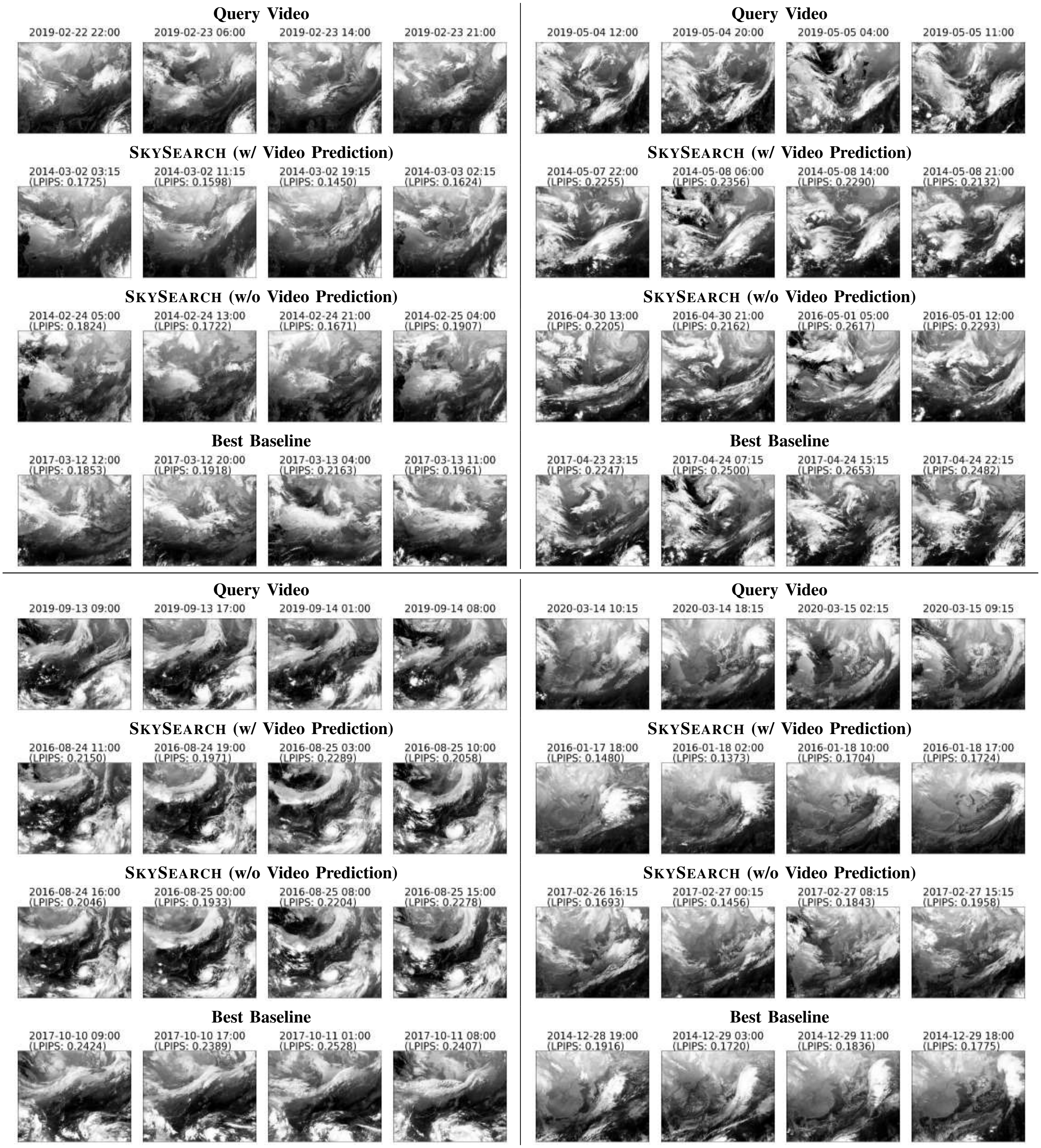
Query Video
2019-02-22 22:00
2019-02-23 06:00
2019-02-23 14:00
2019-02-23 21:00
SkySearch (w/ Video Prediction)
2014-03-02 03:15 (LPIPS: 0.1725)
2014-03-02 11:15 (LPIPS: 0.1598)
2014-03-02 19:15 (LPIPS: 0.1450)
2014-03-03 02:15 (LPIPS: 0.1624)
SkySearch (w/o Video Prediction)
2014-02-24 05:00 (LPIPS: 0.1824)
2014-02-24 13:00 (LPIPS: 0.1722)
2014-02-24 21:00 (LPIPS: 0.1671)
2014-02-25 04:00 (LPIPS: 0.1907)
Best Baseline
2017-03-12 12:00 (LPIPS: 0.1853)
2017-03-12 20:00 (LPIPS: 0.1918)
2017-03-13 04:00 (LPIPS: 0.2163)
2017-03-13 11:00 (LPIPS: 0.1961)
Query Video
2019-05-04 12:00
2019-05-04 20:00
2019-05-05 04:00
2019-05-05 11:00
SkySearch (w/ Video Prediction)
2014-05-07 22:00 (LPIPS: 0.2255)
2014-05-08 06:00 (LPIPS: 0.2356)
2014-05-08 14:00 (LPIPS: 0.2290)
2014-05-08 21:00 (LPIPS: 0.2132)
SkySearch (w/o Video Prediction)
2016-04-30 13:00 (LPIPS: 0.2205)
2016-04-30 21:00 (LPIPS: 0.2162)
2016-05-01 05:00 (LPIPS: 0.2617)
2016-05-01 12:00 (LPIPS: 0.2293)
Best Baseline
2017-04-23 23:15 (LPIPS: 0.2247)
2017-04-24 07:15 (LPIPS: 0.2500)
2017-04-24 15:15 (LPIPS: 0.2653)
2017-04-24 22:15 (LPIPS: 0.2482)
Query Video
2019-09-13 09:00
2019-09-13 17:00
2019-09-14 01:00
2019-09-14 08:00
SkySearch (w/ Video Prediction)
2016-08-24 11:00 (LPIPS: 0.2150)
2016-08-24 19:00 (LPIPS: 0.1971)
2016-08-25 03:00 (LPIPS: 0.2289)
2016-08-25 10:00 (LPIPS: 0.2058)
SkySearch (w/o Video Prediction)
2016-08-24 16:00 (LPIPS: 0.2046)
2016-08-25 00:00 (LPIPS: 0.1933)
2016-08-25 08:00 (LPIPS: 0.2204)
2016-08-25 15:00 (LPIPS: 0.2278)
Best Baseline
2017-10-10 09:00 (LPIPS: 0.2424)
2017-10-10 17:00 (LPIPS: 0.2389)
2017-10-11 01:00 (LPIPS: 0.2528)
2017-10-11 08:00 (LPIPS: 0.2407)
Query Video
2020-03-14 10:15
2020-03-14 18:15
2020-03-15 02:15
2020-03-15 09:15
SkySearch (w/ Video Prediction)
2016-01-17 18:00 (LPIPS: 0.1480)
2016-01-18 02:00 (LPIPS: 0.1373)
2016-01-18 10:00 (LPIPS: 0.1704)
2016-01-18 17:00 (LPIPS: 0.1724)
SkySearch (w/o Video Prediction)
2017-02-26 16:15 (LPIPS: 0.1693)
2017-02-27 00:15 (LPIPS: 0.1456)
2017-02-27 08:15 (LPIPS: 0.1843)
2017-02-27 15:15 (LPIPS: 0.1958)
Best Baseline
2014-12-28 19:00 (LPIPS: 0.1916)
2014-12-29 03:00 (LPIPS: 0.1720)
2014-12-29 11:00 (LPIPS: 0.1836)
2014-12-29 18:00 (LPIPS: 0.1775)

Fig. 1: Qualitative comparison of search results (1 / 2). SkySearch retrieves satellite videos that are more similar to the query videos, compared to those retrieved by the best-performing baseline (i.e., fine-tuned EfficientNet). Moreover, augmenting queries with video prediction enhances 24-hour (long-term) search accuracy.

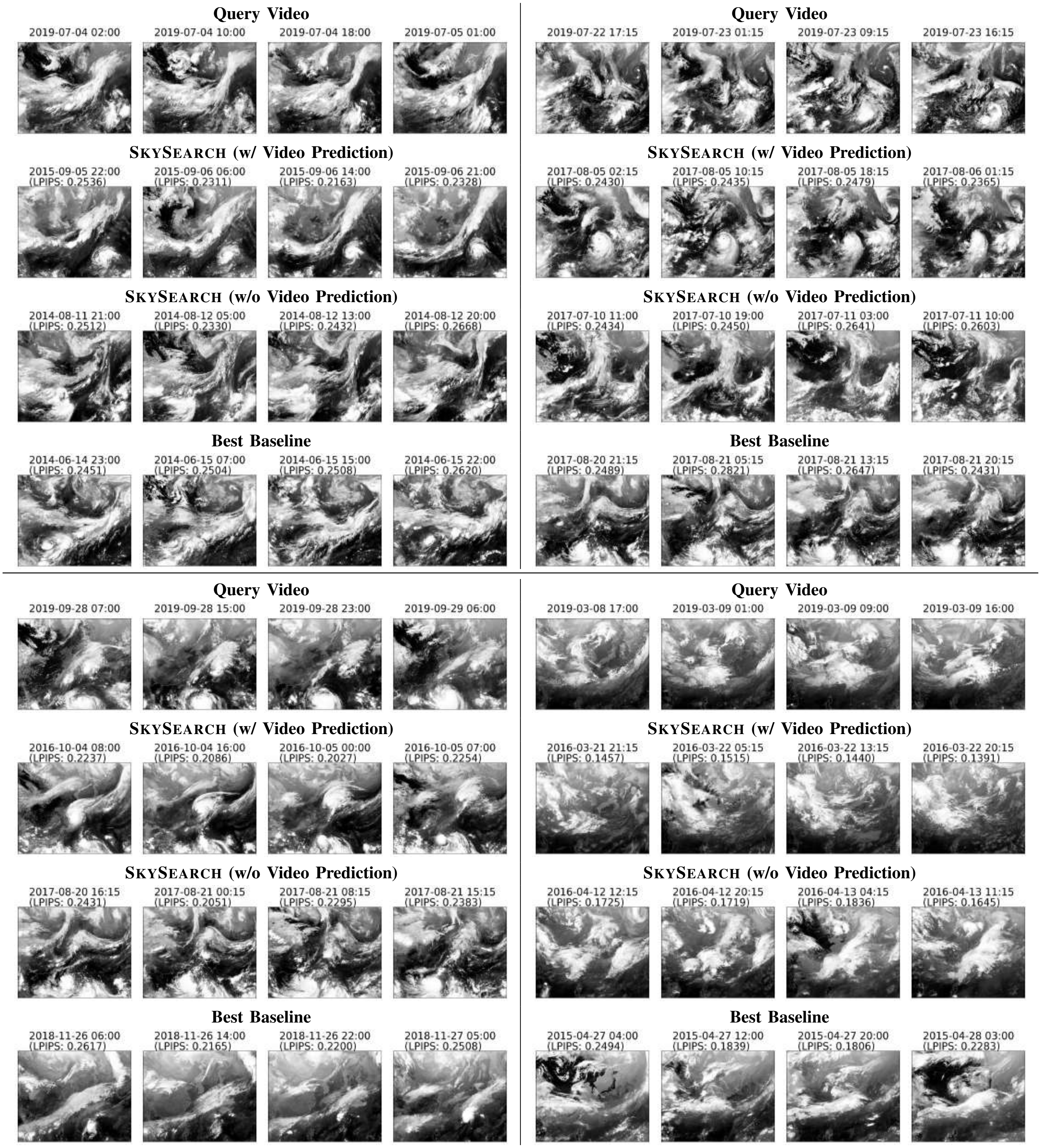
Query Video
2019-07-04 02:00
2019-07-04 10:00
2019-07-04 18:00
2019-07-05 01:00
SkySearch (w/ Video Prediction)
2015-09-05 22:00 (LPIPS: 0.2536)
2015-09-06 06:00 (LPIPS: 0.2311)
2015-09-06 14:00 (LPIPS: 0.2163)
2015-09-06 21:00 (LPIPS: 0.2328)
SkySearch (w/o Video Prediction)
2014-08-11 21:00 (LPIPS: 0.2512)
2014-08-12 05:00 (LPIPS: 0.2330)
2014-08-12 13:00 (LPIPS: 0.2432)
2014-08-12 20:00 (LPIPS: 0.2668)
Best Baseline
2014-06-14 23:00 (LPIPS: 0.2451)
2014-06-15 07:00 (LPIPS: 0.2504)
2014-06-15 15:00 (LPIPS: 0.2508)
2014-06-15 22:00 (LPIPS: 0.2620)
Query Video
2019-07-22 17:15
2019-07-23 01:15
2019-07-23 09:15
2019-07-23 16:15
SkySearch (w/ Video Prediction)
2017-08-05 02:15 (LPIPS: 0.2430)
2017-08-05 10:15 (LPIPS: 0.2435)
2017-08-05 18:15 (LPIPS: 0.2479)
2017-08-06 01:15 (LPIPS: 0.2365)
SkySearch (w/o Video Prediction)
2017-07-10 11:00 (LPIPS: 0.2434)
2017-07-10 19:00 (LPIPS: 0.2450)
2017-07-11 03:00 (LPIPS: 0.2641)
2017-07-11 10:00 (LPIPS: 0.2603)
Best Baseline
2017-08-20 21:15 (LPIPS: 0.2489)
2017-08-21 05:15 (LPIPS: 0.2821)
2017-08-21 13:15 (LPIPS: 0.2647)
2017-08-21 20:15 (LPIPS: 0.2431)
Query Video
2019-09-28 07:00
2019-09-28 15:00
2019-09-28 23:00
2019-09-29 06:00
SkySearch (w/ Video Prediction)
2016-10-04 08:00 (LPIPS: 0.2237)
2016-10-04 16:00 (LPIPS: 0.2086)
2016-10-05 00:00 (LPIPS: 0.2027)
2016-10-05 07:00 (LPIPS: 0.2254)
SkySearch (w/o Video Prediction)
2017-08-20 16:15 (LPIPS: 0.2431)
2017-08-21 00:15 (LPIPS: 0.2051)
2017-08-21 08:15 (LPIPS: 0.2295)
2017-08-21 15:15 (LPIPS: 0.2383)
Best Baseline
2018-11-26 06:00 (LPIPS: 0.2617)
2018-11-26 14:00 (LPIPS: 0.2165)
2018-11-26 22:00 (LPIPS: 0.2200)
2018-11-27 05:00 (LPIPS: 0.2508)
Query Video
2019-03-08 17:00
2019-03-09 01:00
2019-03-09 09:00
2019-03-09 16:00
SkySearch (w/ Video Prediction)
2016-03-21 21:15 (LPIPS: 0.1457)
2016-03-22 05:15 (LPIPS: 0.1515)
2016-03-22 13:15 (LPIPS: 0.1440)
2016-03-22 20:15 (LPIPS: 0.1391)
SkySearch (w/o Video Prediction)
2016-04-12 12:15 (LPIPS: 0.1725)
2016-04-12 20:15 (LPIPS: 0.1719)
2016-04-13 04:15 (LPIPS: 0.1836)
2016-04-13 11:15 (LPIPS: 0.1645)
Best Baseline
2015-04-27 04:00 (LPIPS: 0.2494)
2015-04-27 12:00 (LPIPS: 0.1839)
2015-04-27 20:00 (LPIPS: 0.1806)
2015-04-28 03:00 (LPIPS: 0.2283)

Fig. 2: Qualitative comparison of search results (2 / 2). SkySearch retrieves satellite videos that are more similar to the query videos, compared to those retrieved by the best-performing baseline (i.e., fine-tuned EfficientNet). Moreover, augmenting queries with video prediction enhances 24-hour (long-term) search accuracy.

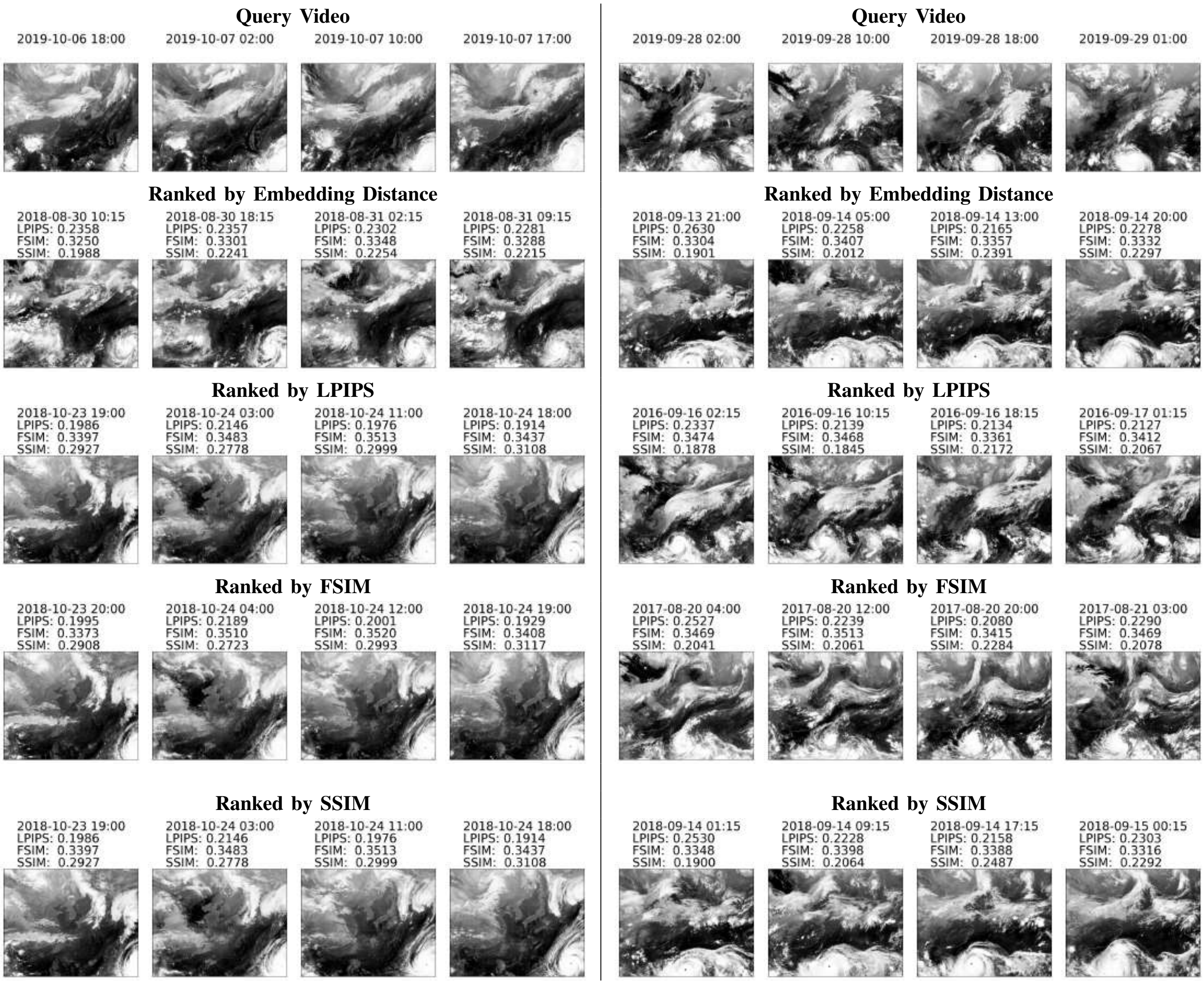
Query Video
2019-10-06 18:00
2019-10-07 02:00
2019-10-07 10:00
2019-10-07 17:00
Ranked by Embedding Distance
2018-08-30 10:15 LPIPS: 0.2358 FSIM: 0.3250 SSIM: 0.1988
2018-08-30 18:15 LPIPS: 0.2357 FSIM: 0.3301 SSIM: 0.2241
2018-08-31 02:15 LPIPS: 0.2302 FSIM: 0.3348 SSIM: 0.2254
2018-08-31 09:15 LPIPS: 0.2281 FSIM: 0.3288 SSIM: 0.2215
Ranked by LPIPS
2018-10-23 19:00 LPIPS: 0.1986 FSIM: 0.3397 SSIM: 0.2927
2018-10-24 03:00 LPIPS: 0.2146 FSIM: 0.3483 SSIM: 0.2778
2018-10-24 11:00 LPIPS: 0.1976 FSIM: 0.3513 SSIM: 0.2999
2018-10-24 18:00 LPIPS: 0.1914 FSIM: 0.3437 SSIM: 0.3108
Ranked by FSIM
2018-10-23 20:00 LPIPS: 0.1995 FSIM: 0.3373 SSIM: 0.2908
2018-10-24 04:00 LPIPS: 0.2189 FSIM: 0.3510 SSIM: 0.2723
2018-10-24 12:00 LPIPS: 0.2001 FSIM: 0.3520 SSIM: 0.2993
2018-10-24 19:00 LPIPS: 0.1929 FSIM: 0.3408 SSIM: 0.3117
Ranked by SSIM
2018-10-23 19:00 LPIPS: 0.1986 FSIM: 0.3397 SSIM: 0.2927
2018-10-24 03:00 LPIPS: 0.2146 FSIM: 0.3483 SSIM: 0.2778
2018-10-24 11:00 LPIPS: 0.1976 FSIM: 0.3513 SSIM: 0.2999
2018-10-24 18:00 LPIPS: 0.1914 FSIM: 0.3437 SSIM: 0.3108
Query Video
2019-09-28 02:00
2019-09-28 10:00
2019-09-28 18:00
2019-09-29 01:00
Ranked by Embedding Distance
2018-09-13 21:00 LPIPS: 0.2630 FSIM: 0.3304 SSIM: 0.1901
2018-09-14 05:00 LPIPS: 0.2258 FSIM: 0.3407 SSIM: 0.2012
2018-09-14 13:00 LPIPS: 0.2165 FSIM: 0.3357 SSIM: 0.2391
2018-09-14 20:00 LPIPS: 0.2278 FSIM: 0.3332 SSIM: 0.2297
Ranked by LPIPS
2016-09-16 02:15 LPIPS: 0.2337 FSIM: 0.3474 SSIM: 0.1878
2016-09-16 10:15 LPIPS: 0.2139 FSIM: 0.3468 SSIM: 0.1845
2016-09-16 18:15 LPIPS: 0.2134 FSIM: 0.3361 SSIM: 0.2172
2016-09-17 01:15 LPIPS: 0.2127 FSIM: 0.3412 SSIM: 0.2067
Ranked by FSIM
2017-08-20 04:00 LPIPS: 0.2527 FSIM: 0.3469 SSIM: 0.2041
2017-08-20 12:00 LPIPS: 0.2239 FSIM: 0.3513 SSIM: 0.2061
2017-08-20 20:00 LPIPS: 0.2080 FSIM: 0.3415 SSIM: 0.2284
2017-08-21 03:00 LPIPS: 0.2290 FSIM: 0.3469 SSIM: 0.2078
Ranked by SSIM
2018-09-14 01:15 LPIPS: 0.2530 FSIM: 0.3348 SSIM: 0.1900
2018-09-14 09:15 LPIPS: 0.2228 FSIM: 0.3398 SSIM: 0.2064
2018-09-14 17:15 LPIPS: 0.2158 FSIM: 0.3388 SSIM: 0.2487
2018-09-15 00:15 LPIPS: 0.2303 FSIM: 0.3316 SSIM: 0.2292

Fig. 3: Results of Alternative Ranking Methods (1 / 2) SkySearch ranks the most similar videos from the candidates retrieved by the candidate search module using embedding distances, LPIPS, SSIM, or FSIM. While all retrieved videos are visually similar to the query, image-based measures such as LPIPS, SSIM, and FSIM can provide further refinement at the cost of higher computational complexity.

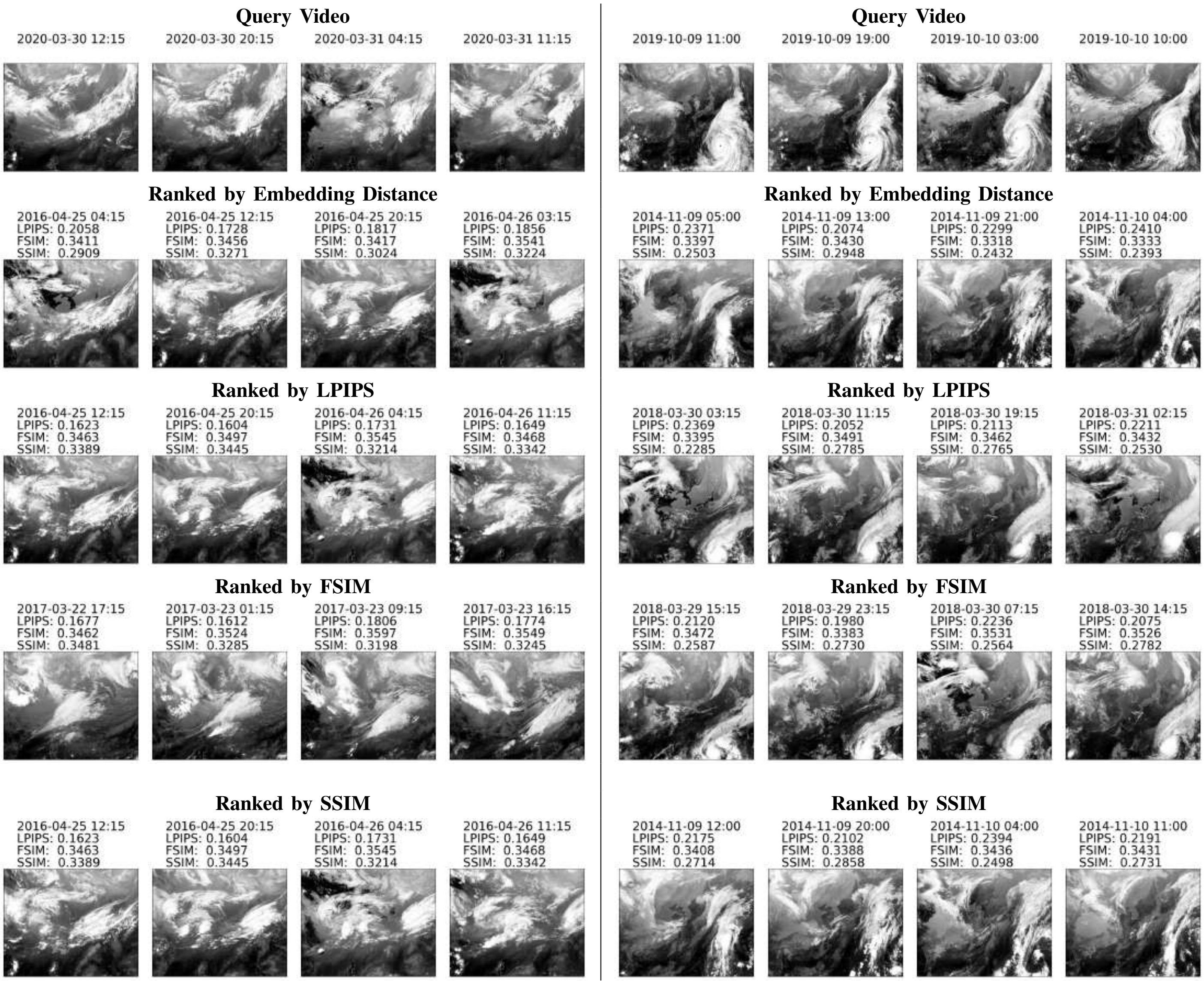
Query Video
2020-03-30 12:15
2020-03-30 20:15
2020-03-31 04:15
2020-03-31 11:15
Ranked by Embedding Distance
2016-04-25 04:15 LPIPS: 0.2058 FSIM: 0.3411 SSIM: 0.2909
2016-04-25 12:15 LPIPS: 0.1728 FSIM: 0.3456 SSIM: 0.3271
2016-04-25 20:15 LPIPS: 0.1817 FSIM: 0.3417 SSIM: 0.3024
2016-04-26 03:15 LPIPS: 0.1856 FSIM: 0.3541 SSIM: 0.3224
Ranked by LPIPS
2016-04-25 12:15 LPIPS: 0.1623 FSIM: 0.3463 SSIM: 0.3389
2016-04-25 20:15 LPIPS: 0.1604 FSIM: 0.3497 SSIM: 0.3445
2016-04-26 04:15 LPIPS: 0.1731 FSIM: 0.3545 SSIM: 0.3214
2016-04-26 11:15 LPIPS: 0.1649 FSIM: 0.3468 SSIM: 0.3342
Ranked by FSIM
2017-03-22 17:15 LPIPS: 0.1677 FSIM: 0.3462 SSIM: 0.3481
2017-03-23 01:15 LPIPS: 0.1612 FSIM: 0.3524 SSIM: 0.3285
2017-03-23 09:15 LPIPS: 0.1806 FSIM: 0.3597 SSIM: 0.3198
2017-03-23 16:15 LPIPS: 0.1774 FSIM: 0.3549 SSIM: 0.3245
Ranked by SSIM
2016-04-25 12:15 LPIPS: 0.1623 FSIM: 0.3463 SSIM: 0.3389
2016-04-25 20:15 LPIPS: 0.1604 FSIM: 0.3497 SSIM: 0.3445
2016-04-26 04:15 LPIPS: 0.1731 FSIM: 0.3545 SSIM: 0.3214
2016-04-26 11:15 LPIPS: 0.1649 FSIM: 0.3468 SSIM: 0.3342
Query Video
2019-10-09 11:00
2019-10-09 19:00
2019-10-10 03:00
2019-10-10 10:00
Ranked by Embedding Distance
2014-11-09 05:00 LPIPS: 0.2371 FSIM: 0.3397 SSIM: 0.2503
2014-11-09 13:00 LPIPS: 0.2074 FSIM: 0.3430 SSIM: 0.2948
2014-11-09 21:00 LPIPS: 0.2299 FSIM: 0.3318 SSIM: 0.2432
2014-11-10 04:00 LPIPS: 0.2410 FSIM: 0.3333 SSIM: 0.2393
Ranked by LPIPS
2018-03-30 03:15 LPIPS: 0.2369 FSIM: 0.3395 SSIM: 0.2285
2018-03-30 11:15 LPIPS: 0.2052 FSIM: 0.3491 SSIM: 0.2785
2018-03-30 19:15 LPIPS: 0.2113 FSIM: 0.3462 SSIM: 0.2765
2018-03-31 02:15 LPIPS: 0.2211 FSIM: 0.3432 SSIM: 0.2530
Ranked by FSIM
2018-03-29 15:15 LPIPS: 0.2120 FSIM: 0.3472 SSIM: 0.2587
2018-03-29 23:15 LPIPS: 0.1980 FSIM: 0.3383 SSIM: 0.2730
2018-03-30 07:15 LPIPS: 0.2236 FSIM: 0.3531 SSIM: 0.2564
2018-03-30 14:15 LPIPS: 0.2075 FSIM: 0.3526 SSIM: 0.2782
Ranked by SSIM
2014-11-09 12:00 LPIPS: 0.2175 FSIM: 0.3408 SSIM: 0.2714
2014-11-09 20:00 LPIPS: 0.2102 FSIM: 0.3388 SSIM: 0.2858
2014-11-10 04:00 LPIPS: 0.2394 FSIM: 0.3436 SSIM: 0.2498
2014-11-10 11:00 LPIPS: 0.2191 FSIM: 0.3431 SSIM: 0.2731

Fig. 4: Results of Alternative Ranking Methods (2 / 2) SkySearch ranks the most similar videos from the candidates retrieved by the candidate search module using embedding distances, LPIPS, SSIM, or FSIM. While all retrieved videos are visually similar to the query, image-based measures such as LPIPS, SSIM, and FSIM can provide further refinement at the cost of higher computational complexity.

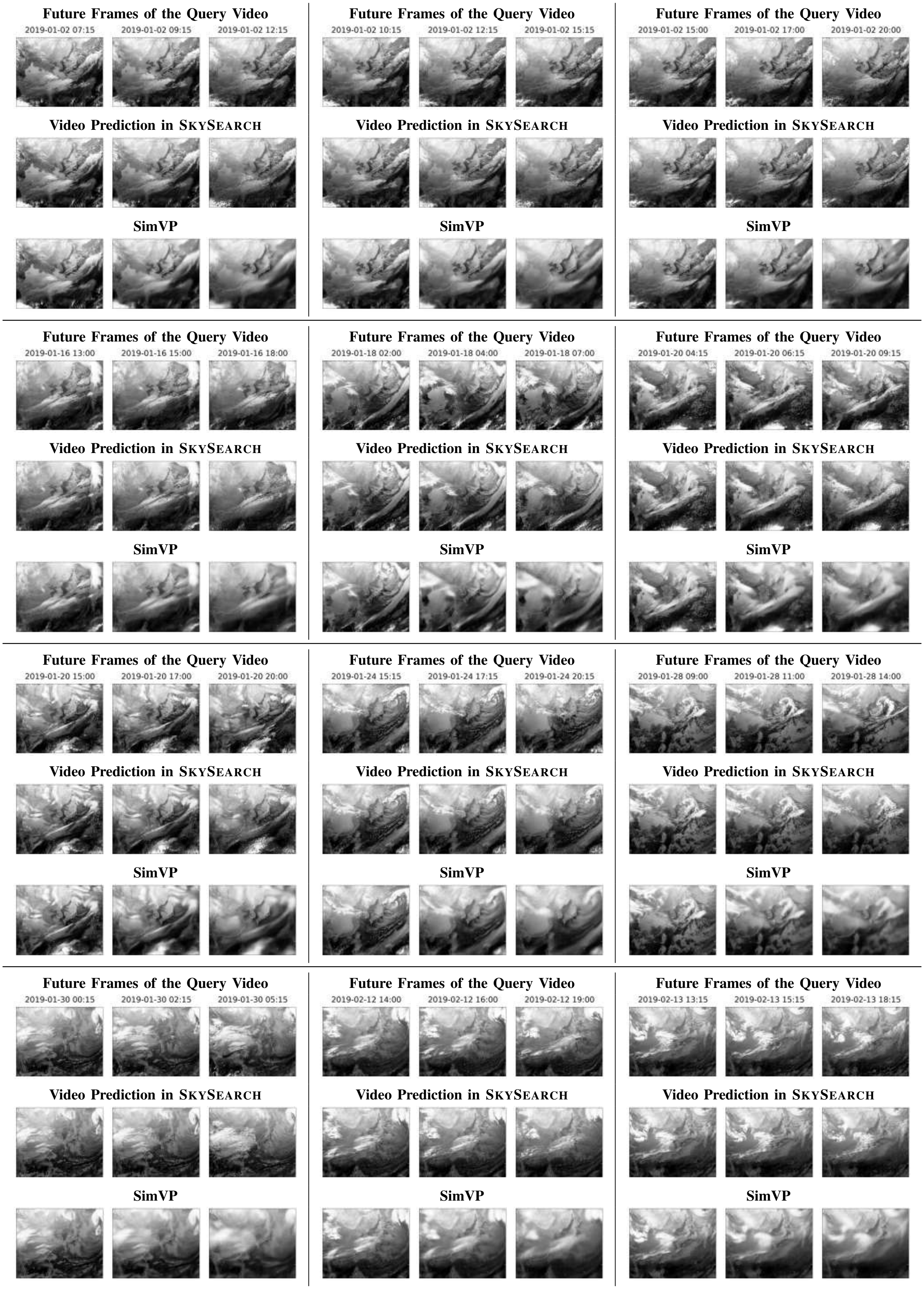
Future Frames of the Query Video
2019-01-02 07:15
2019-01-02 09:15
2019-01-02 12:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-02 10:15
2019-01-02 12:15
2019-01-02 15:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-02 15:00
2019-01-02 17:00
2019-01-02 20:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-16 13:00
2019-01-16 15:00
2019-01-16 18:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-18 02:00
2019-01-18 04:00
2019-01-18 07:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-20 04:15
2019-01-20 06:15
2019-01-20 09:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-20 15:00
2019-01-20 17:00
2019-01-20 20:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-24 15:15
2019-01-24 17:15
2019-01-24 20:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-28 09:00
2019-01-28 11:00
2019-01-28 14:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-01-30 00:15
2019-01-30 02:15
2019-01-30 05:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-02-12 14:00
2019-02-12 16:00
2019-02-12 19:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-02-13 13:15
2019-02-13 15:15
2019-02-13 18:15
Video Prediction in SkySearch
SimVP

Fig. 5: Video prediction results **(1 / 2)**. SkySearch employs adversarial learning for video prediction, leading to more accurate and sharper future predictions of the query video compared to SimVP [1].

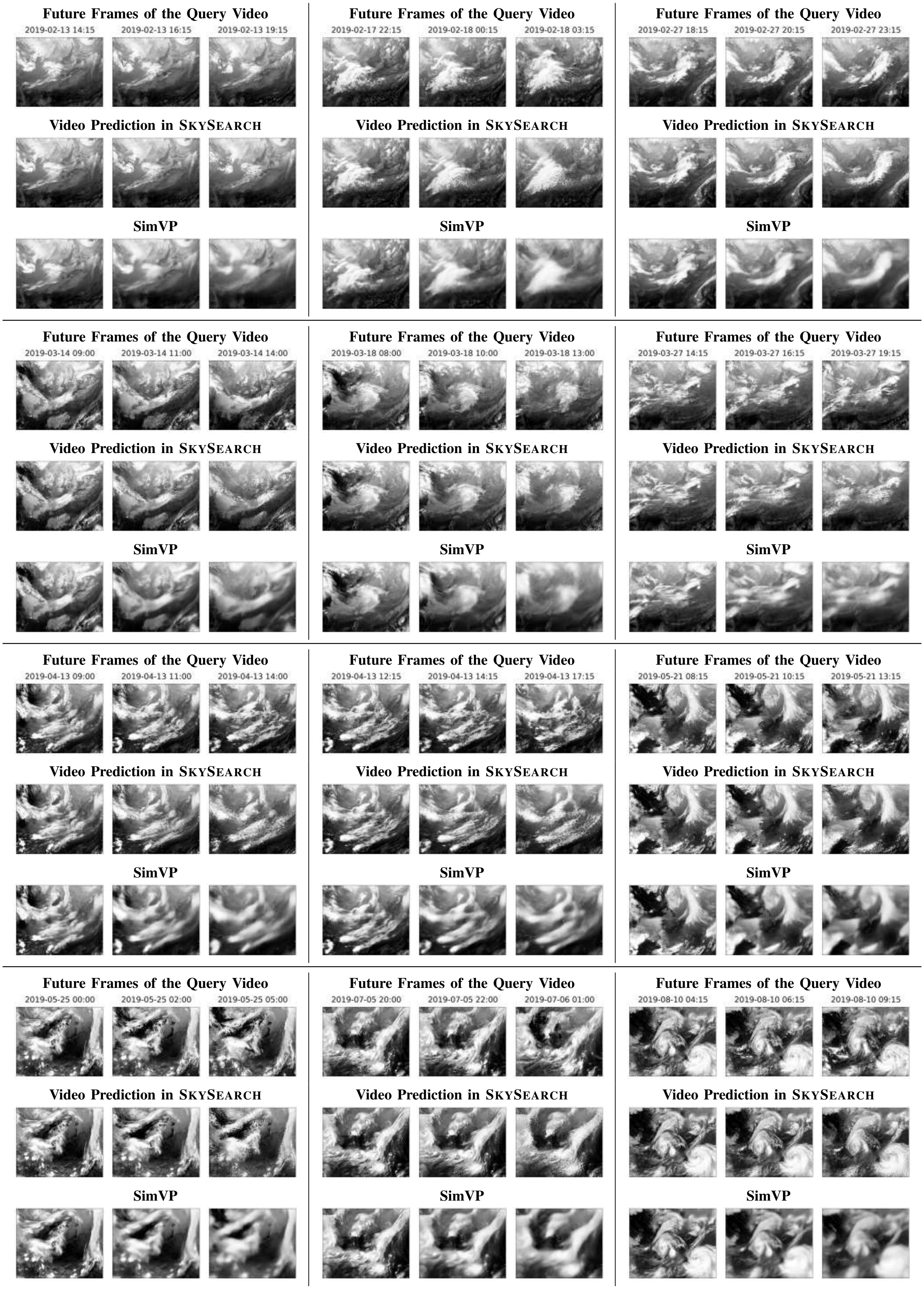
Future Frames of the Query Video
2019-02-13 14:15
2019-02-13 16:15
2019-02-13 19:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-02-17 22:15
2019-02-18 00:15
2019-02-18 03:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-02-27 18:15
2019-02-27 20:15
2019-02-27 23:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-03-14 09:00
2019-03-14 11:00
2019-03-14 14:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-03-18 08:00
2019-03-18 10:00
2019-03-18 13:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-03-27 14:15
2019-03-27 16:15
2019-03-27 19:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-04-13 09:00
2019-04-13 11:00
2019-04-13 14:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-04-13 12:15
2019-04-13 14:15
2019-04-13 17:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-05-21 08:15
2019-05-21 10:15
2019-05-21 13:15
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-05-25 00:00
2019-05-25 02:00
2019-05-25 05:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-07-05 20:00
2019-07-05 22:00
2019-07-06 01:00
Video Prediction in SkySearch
SimVP
Future Frames of the Query Video
2019-08-10 04:15
2019-08-10 06:15
2019-08-10 09:15
Video Prediction in SkySearch
SimVP

Fig. 6: Video prediction results **(2 / 2)**. SkySearch employs adversarial learning for video prediction, leading to more accurate and sharper future predictions of the query video compared to SimVP [1].