Visualization-Driven Illumination for Density Plots

– Supplementary Material –

Xin Chen, Yunhai Wang, Huaiwei Bao, Kecheng Lu, Jaemin Jo, Chi-Wing Fu, Jean-Daniel Fekete

This supplemental material file includes five additional results for our submitted paper: (i) detailed comparisons of the automatic lighting setup vs. the default setting across 10 tested datasets, (ii) more illumination results under different settings, (iii) engagement checks with accompanying questions and answers, (iv) a case study comparing VIDP with existing density plot enhancement techniques, and (v) the images of density plots used in the experiment and the associated results of each task.

Comparison Between Automatic and Default Lighting Setup. In our submitted paper, we conducted a comparative analysis of automatic and default lighting setups using the *Hertzsprung-Russell diagram* dataset. Here, we present the overall variances for ten tested datasets under both lighting setups in Table 1. The results indicate that our automatic lighting setup outperforms the default configuration (azimuth=120°, elevation=60°) in 9 out of 10 tested datasets. Notably, even in the one dataset where the automatic setup performed slightly worse, the difference was minimal. For details on the calculation process, the code can be found in the file "verifying_automatic_lighting_setup.ipynb" within the "experiment.zip".

Dataset	Credit card fraud	Diabetes	Facial expressions	Person activity	Satimage	Synthesis1	Synthesis2	Synthesis3	Synthesis4	Synthesis5
Automatic lighting	0.023083	0.043872	0.018502	0.066057	0.007121	0.014273	0.045813	0.063650	0.078243	0.132787
Default lighting	0.022966	0.043474	0.018495	0.065749	0.007073	0.014240	0.045703	0.063604	0.078143	0.132795

Table 1: The overall variances for the ten tested datasets under automatic and default lighting setups. Higher variances are highlighted with a red background.

Illumination Results. In our submitted paper, we conduct a parameter analysis employing a light background alongside a perceptually uniform colormap, Magma [5]. Here, we provide additional illumination results under the dark versions of a perceptually uniform colormap Plasma (Figure 1) and a single-hue sequential colormap Reds (Figure 2). Note that, we use positive ϕ to lighten structures in the dark background. In addition, we provide illumination results on the real-world datasets used in our experiment and find that $\eta = 5, \phi = -25$ achieves a good balance between the visibility of outliers and color distortion.

Engagement Checks. In our user study, we included an engagement check trial for each task. Figure 8 displays the screenshots of these trials along with the corresponding questions, where the answers are readily apparent.

Additional Case Study: Diamonds. We collected this classic dataset from Kaggle [6], which contains the records of 53,940 diamonds. By mapping the length and price properties to the x and y axes and encoding the result using the Plasma color map [5], we obtained the scatterplot in Figure 9a. It clearly shows the outliers in the red box but obscures high-density regions due to overplotting. As shown in Figure 9b-9f, details of the highest-density region (green box) can be discerned in all techniques except SPP. However, Figure 9d shows IDP would produce shading on one side of the ridge and interfere with the density estimation, while Figure 9e shows SUP would introduce artificial white colors, preventing to perceive absolute density values accurately. In contrast, VIDP changed the luminance but did not result in conspicuous artifacts (Figure 9f). As for the outliers in the red box, we can find them in Figure 9c,9e,9f, but not in Figure 9b,9d. Nevertheless, SSP and SUP paid the price of color ambiguity [7] to reveal outliers, while our VIDP preserved the relative density in terms of color. Overall, our VIDP technique is still effective even in a dark background.

Density Plots in the Evaluation. In the end, we present the density plot images used in our controlled user study of ten datasets, comprising five real-world and five synthetic ones. The abbreviations for three existing techniques and our technique used to generate these images are as follows:

- *X. Chen, H. Bao are with Shandong University, CN. E-mail: {chenxin199634,bhuaiwei}@gmail.com.*
- *K. Lu is with Renmin University of China, CN. E-mail: lukecheng0407@gmail.com*
- *J. Jo is with Sungkyunkwan University, KR. E-mail: jmjo@skku.edu.*
- *C.-W. Fu is with the Chinese University of Hong Kong, Hong Kong, CN. Email: cwfu@cse.cuhk.edu.hk*
- *J.-D. Fekete is with University Paris-Saclay, CNRS, Inria, LISN, FR. E-mail: Jean-Daniel.Fekete@inria.fr*
- *Y. Wang is with Renmin University of China, CN, and is the corresponding author. E-mail: wang.yh@ruc.edu.cn*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

- **CDP**: Continuous density plot
- **IDP**: Illuminated Density Plot [8]
- **SUP**: Sunspot plot [10]
- **VIDP**: Visualization-driven Illuminated Density Plot (Ours)

In addition, the average error rates of individual datasets for the three tasks (*density comparison* (T1), *density estimation* (T2), and *outlier identification* (T3)) are shown below the corresponding images. The technique with the lowest error rate among the four is highlighted in bold. As we can see, VIDP performs best in 9, 5, and 6 of the 10 datasets for T1, T2, and T3, respectively. When it is not ranked first, it consistently ranks second. Such results indicate the robustness of our approach across different datasets.

The full evaluation results, including the user study website and the analysis code, are available at https://osf.io/5xpsw/?view_only=0445046dad574d4a90d7138e94547ada.

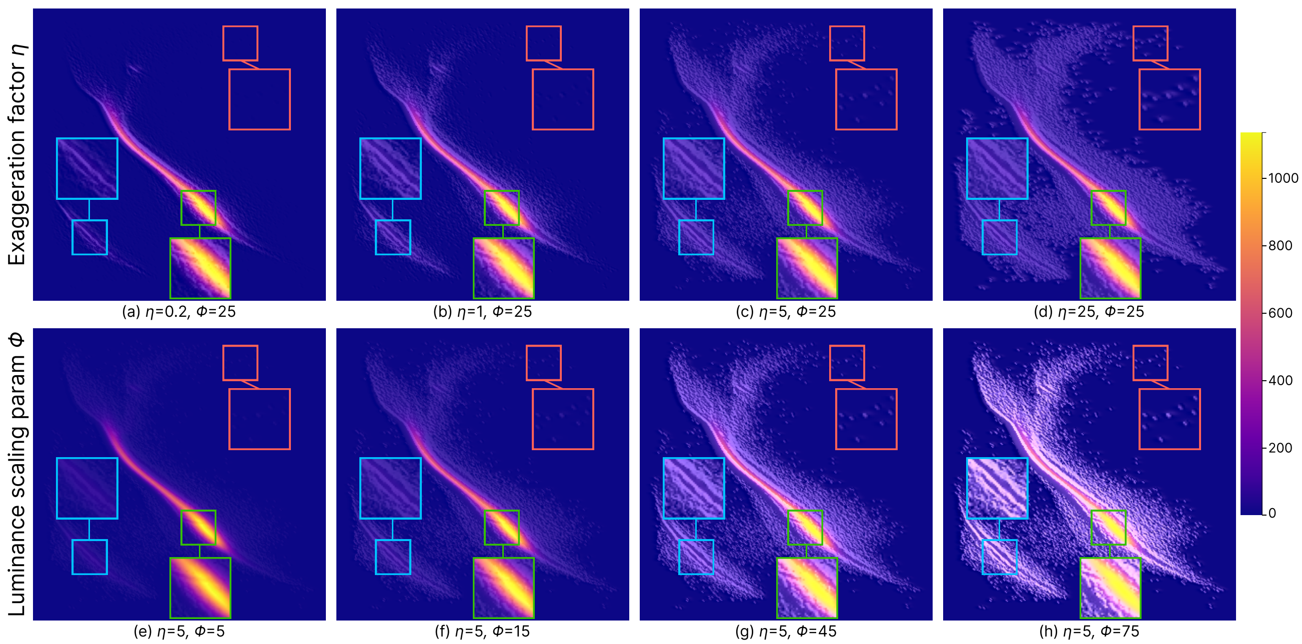


Fig. 1: Parameter analysis on the *Hertzsprung-Russell diagram* dataset [1] using the perceptually uniform colormap, Plasma.

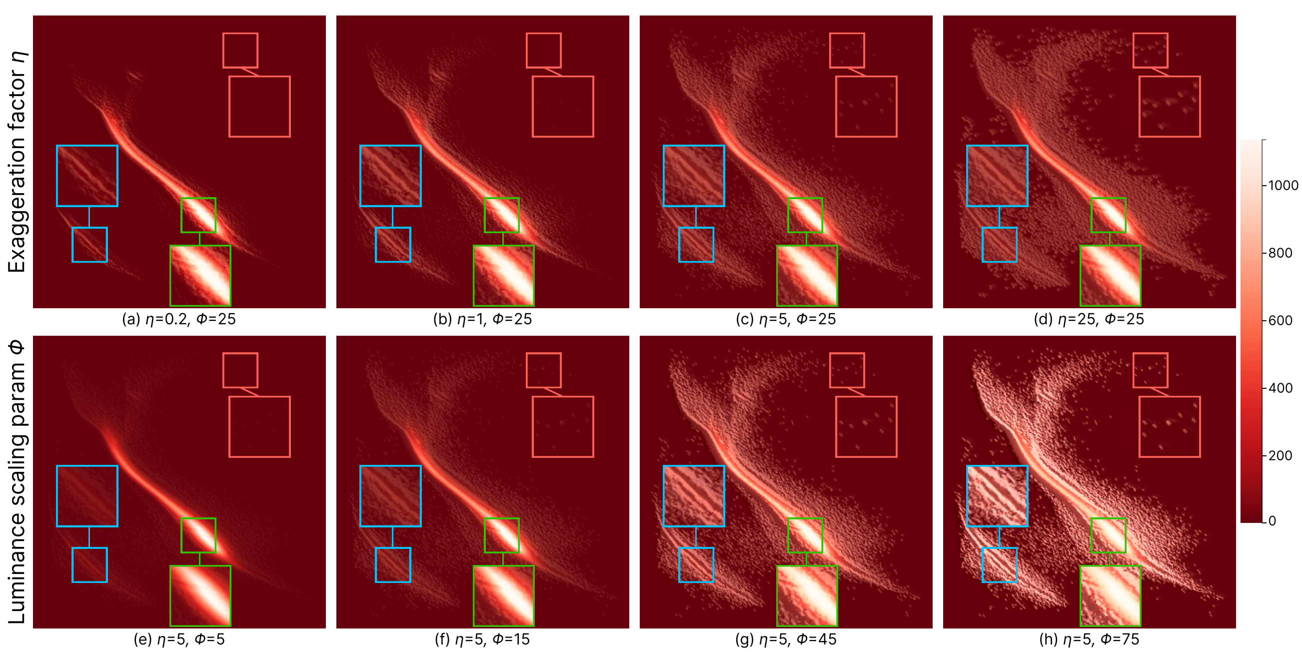


Fig. 2: Parameter analysis on the *Hertzsprung-Russell diagram* dataset [1] using the single-hue sequential colormap, Reds.

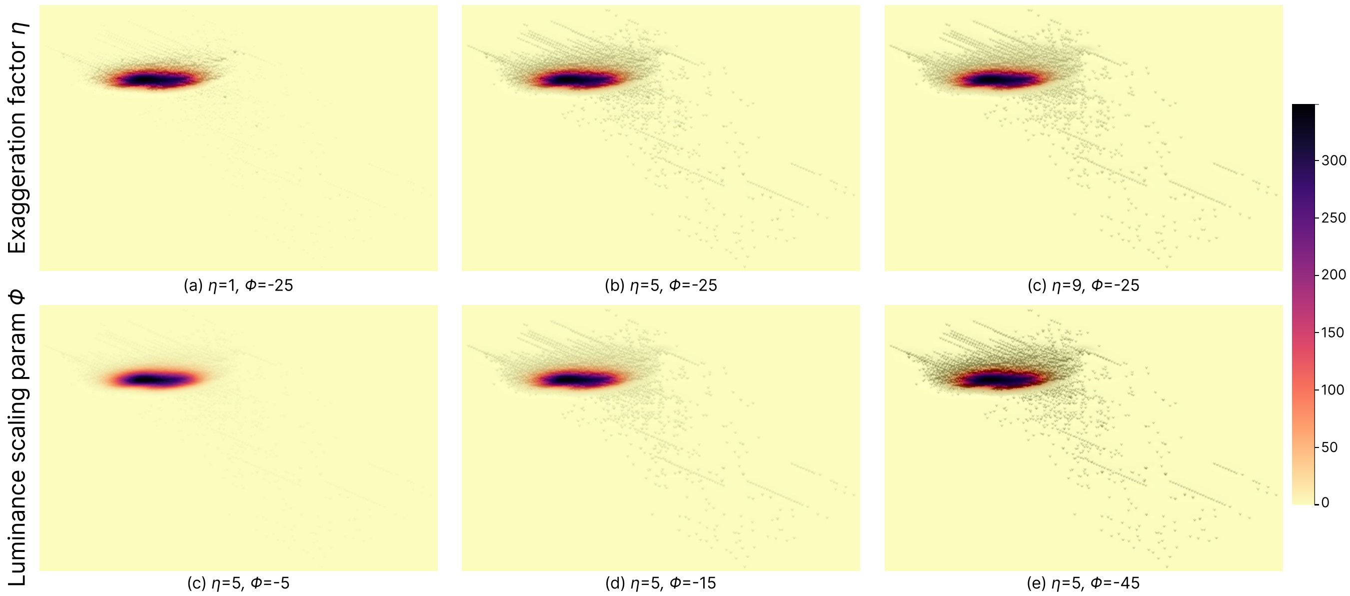


Fig. 3: The effects of the parameters on the *Credit card fraud* dataset [3].

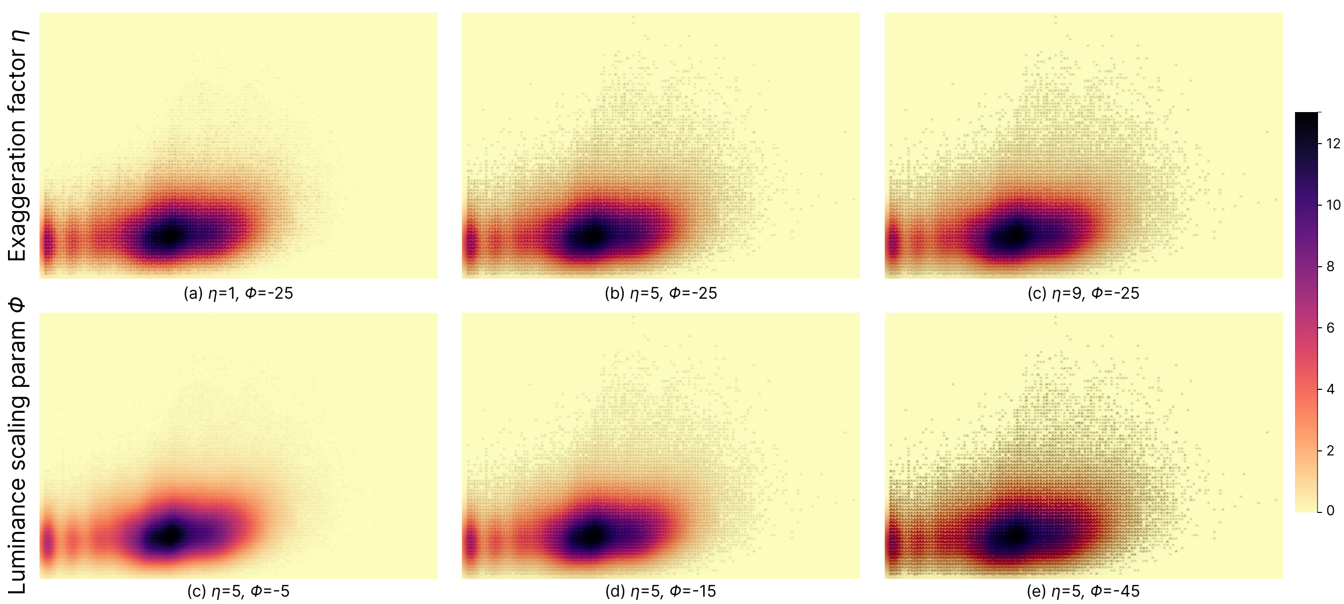


Fig. 4: The effects of the parameters on the *Diabetes* dataset [2].

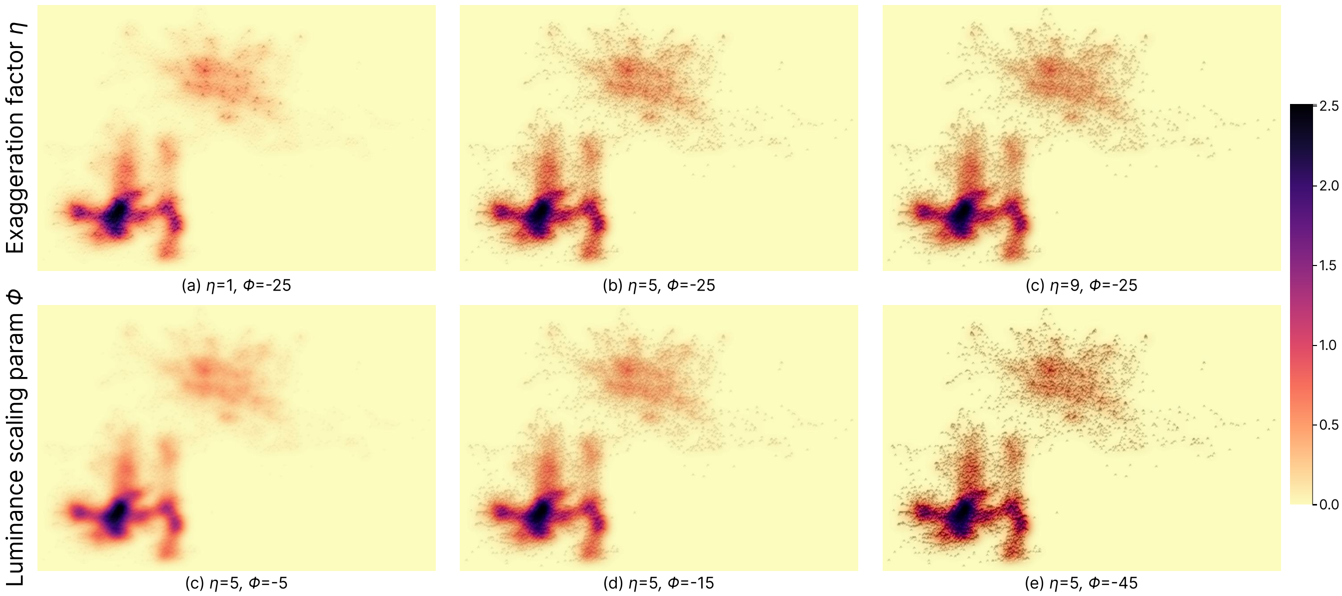


Fig. 5: The effects of the parameters on the *Facial expressions* dataset [4].

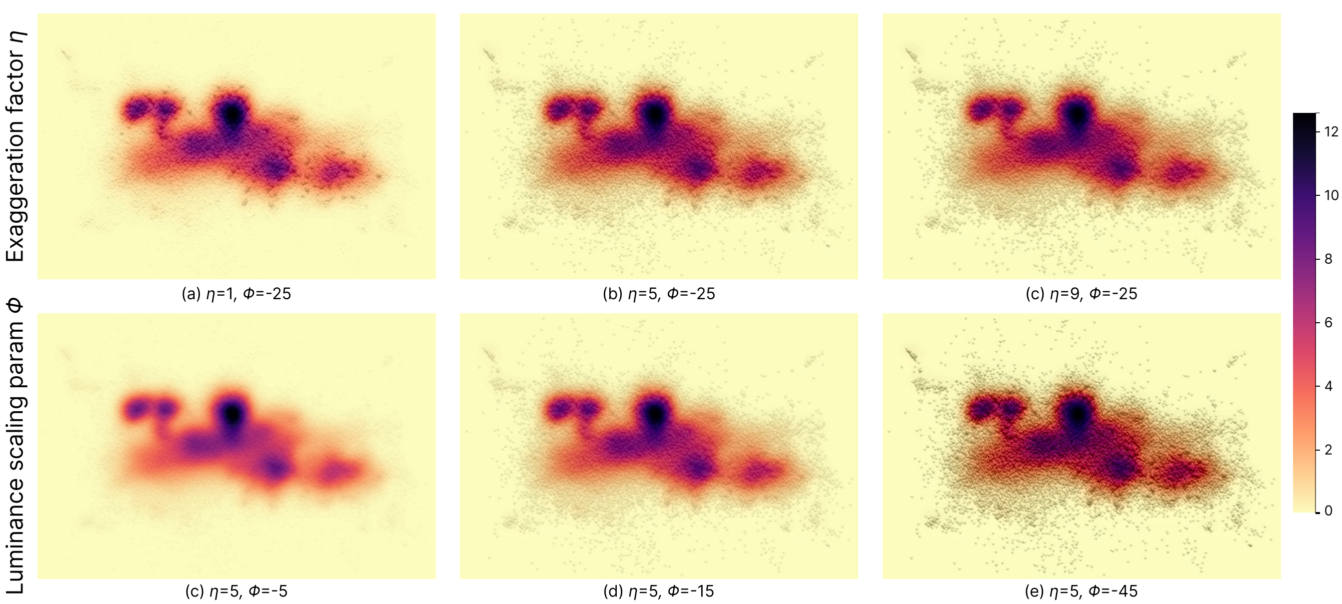


Fig. 6: The effects of the parameters on the *Person Activity* dataset [11].

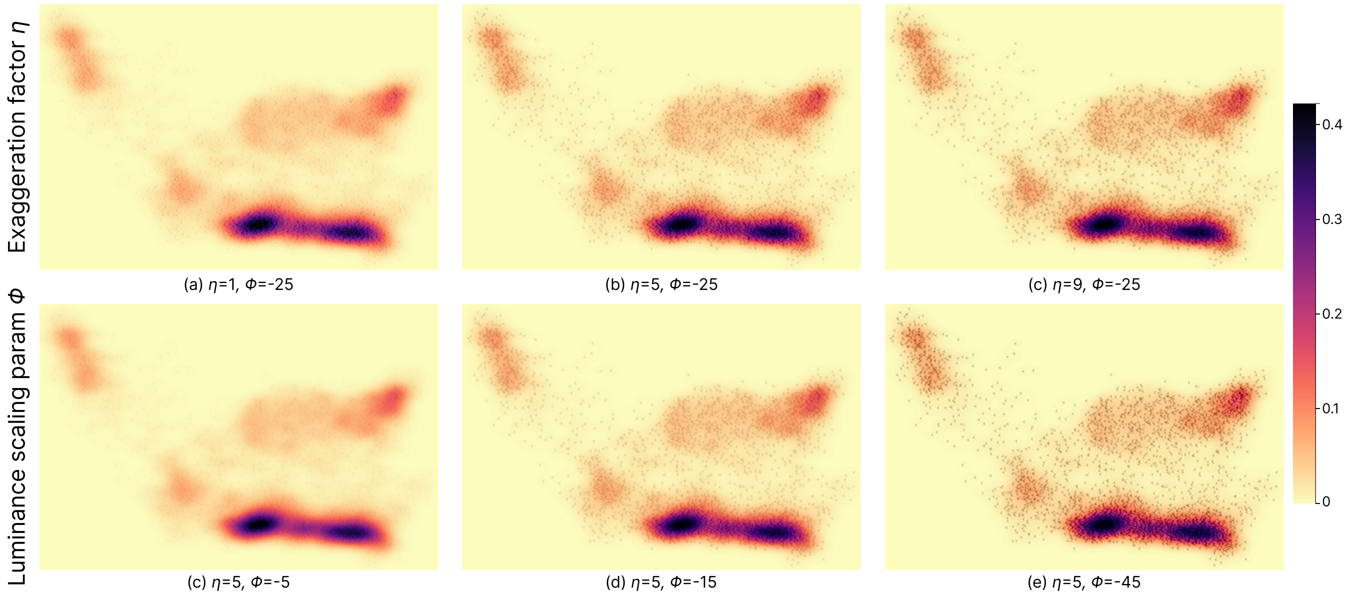


Fig. 7: The effects of the parameters on the *Satimage* dataset [9].

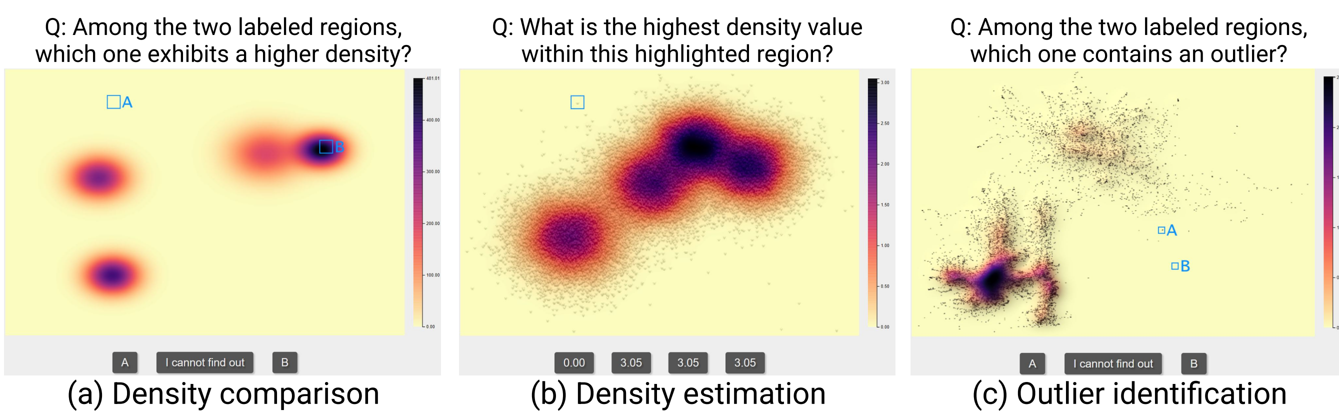


Fig. 8: Screenshots of engagement check trials for the three tasks in our user study and the corresponding questions we asked participants.

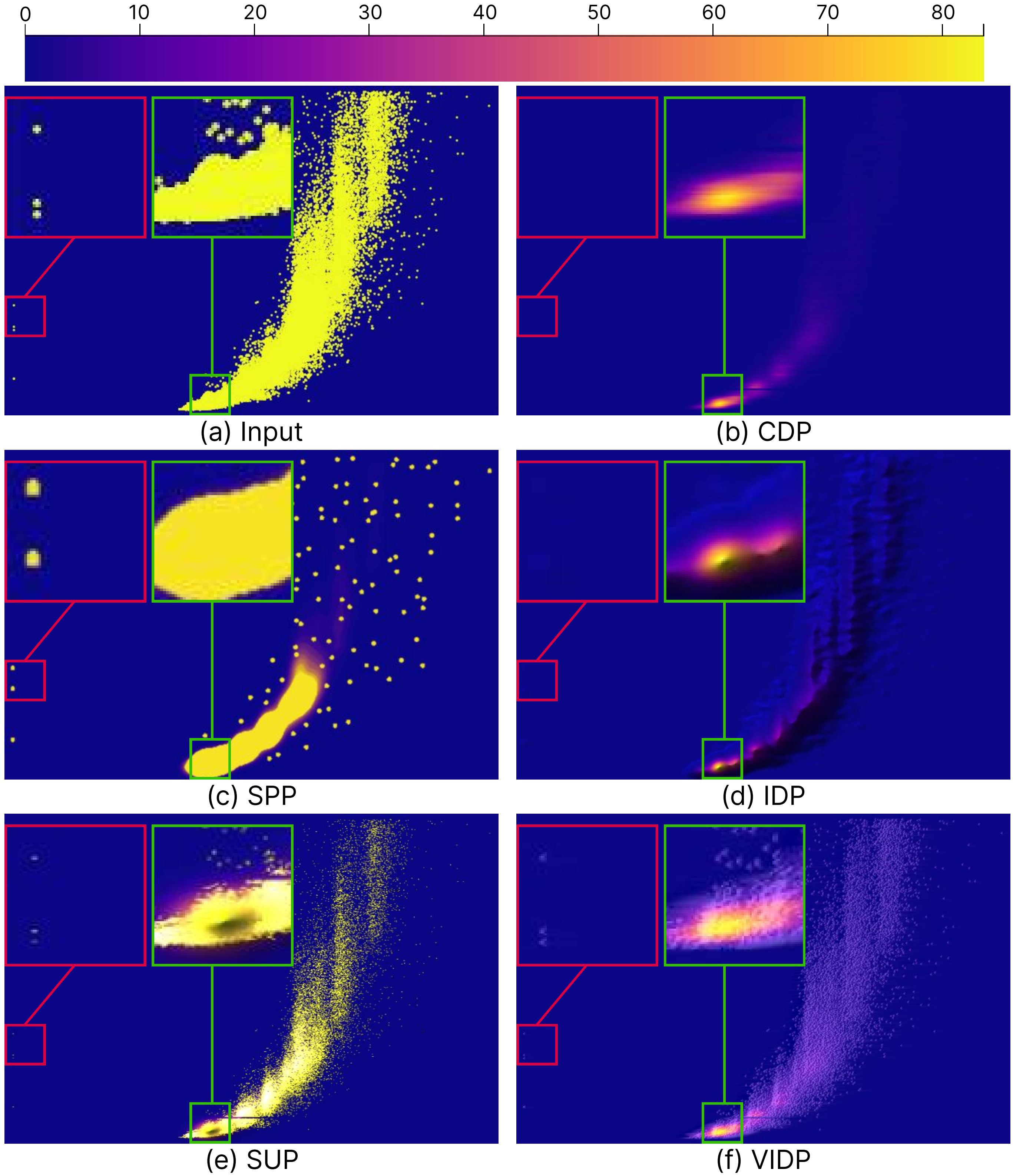
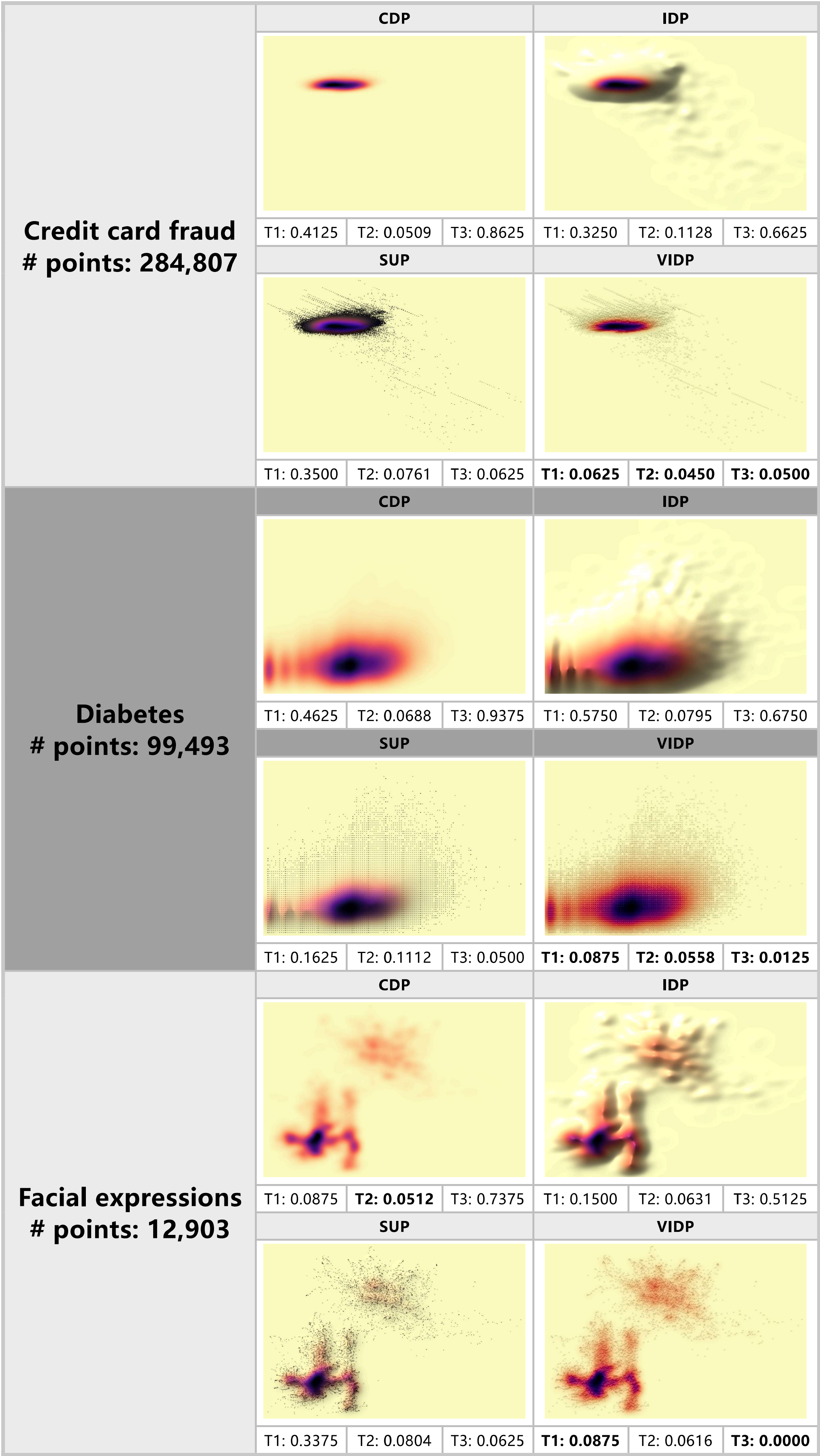
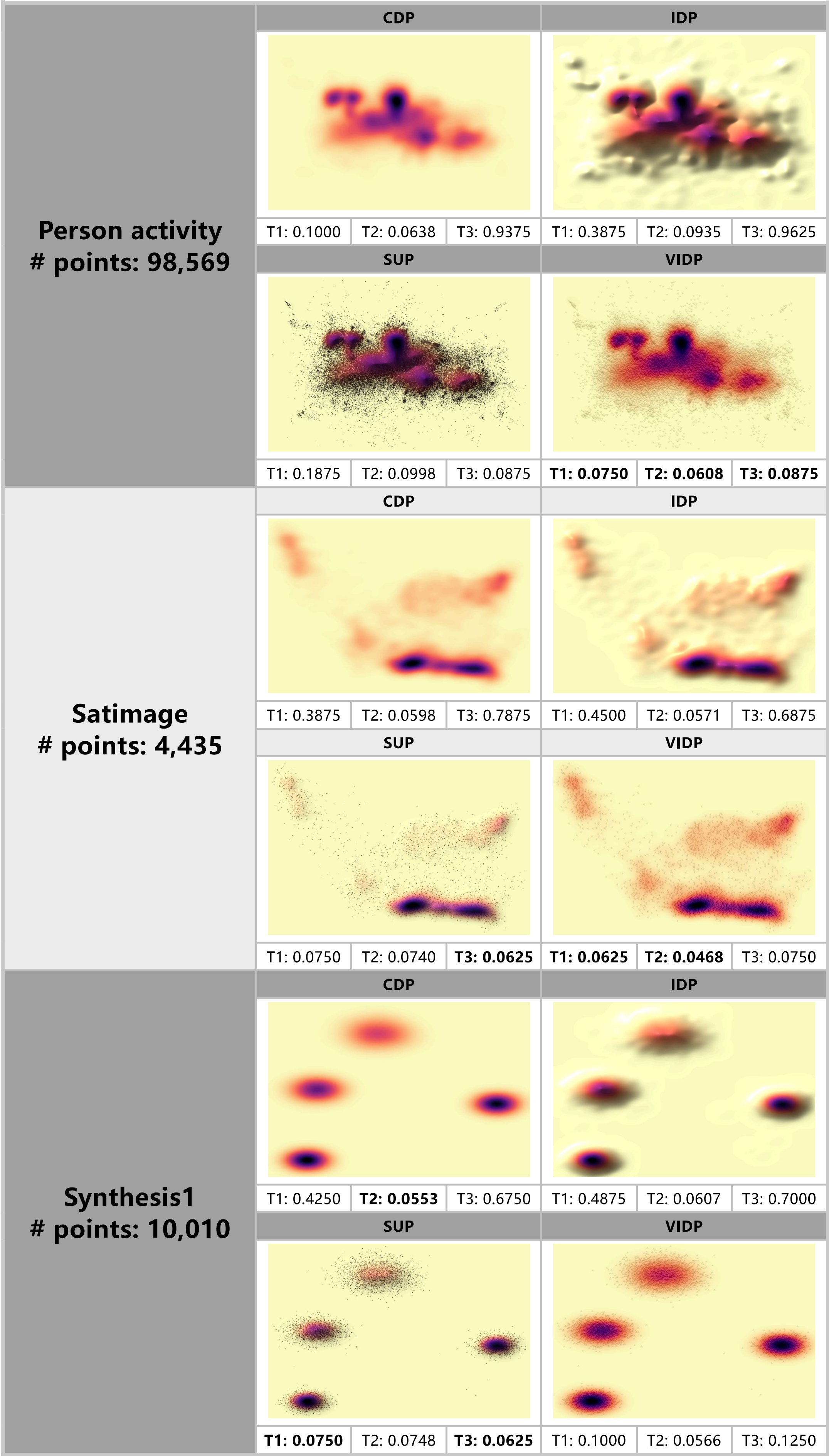


Fig. 9: Density plots of the classic *diamonds* dataset [6] in a dark background. (a) The original scatterplot clearly shows outliers (red box) but obscures the highest-density region (green box). (b-f) Results generated by continuous density plot (CDP), Splatterplot (SSP), illuminated density plot (IDP), sunspot plot (SUP), and our visualization-driven illuminated density plot (VIDP).

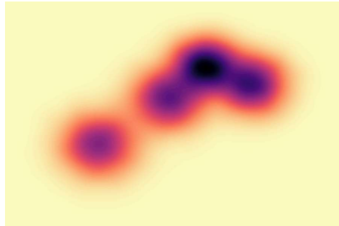

Credit card fraud
points: 284,807
CDP
T1: 0.4125
T2: 0.0509
T3: 0.8625
IDP
T1: 0.3250
T2: 0.1128
T3: 0.6625
SUP
T1: 0.3500
T2: 0.0761
T3: 0.0625
VIDP
T1: 0.0625
T2: 0.0450
T3: 0.0500
Diabetes
points: 99,493
CDP
T1: 0.4625
T2: 0.0688
T3: 0.9375
IDP
T1: 0.5750
T2: 0.0795
T3: 0.6750
SUP
T1: 0.1625
T2: 0.1112
T3: 0.0500
VIDP
T1: 0.0875
T2: 0.0558
T3: 0.0125
Facial expressions
points: 12,903
CDP
T1: 0.0875
T2: 0.0512
T3: 0.7375
IDP
T1: 0.1500
T2: 0.0631
T3: 0.5125
SUP
T1: 0.3375
T2: 0.0804
T3: 0.0625
VIDP
T1: 0.0875
T2: 0.0616
T3: 0.0000



Person activity
points: 98,569
CDP
T1: 0.1000
T2: 0.0638
T3: 0.9375
IDP
T1: 0.3875
T2: 0.0935
T3: 0.9625
SUP
T1: 0.1875
T2: 0.0998
T3: 0.0875
VIDP
T1: 0.0750
T2: 0.0608
T3: 0.0875
Satimage
points: 4,435
CDP
T1: 0.3875
T2: 0.0598
T3: 0.7875
IDP
T1: 0.4500
T2: 0.0571
T3: 0.6875
SUP
T1: 0.0750
T2: 0.0740
T3: 0.0625
VIDP
T1: 0.0625
T2: 0.0468
T3: 0.0750
Synthesis1
points: 10,010
CDP
T1: 0.4250
T2: 0.0553
T3: 0.6750
IDP
T1: 0.4875
T2: 0.0607
T3: 0.7000
SUP
T1: 0.0750
T2: 0.0748
T3: 0.0625
VIDP
T1: 0.1000
T2: 0.0566
T3: 0.1250

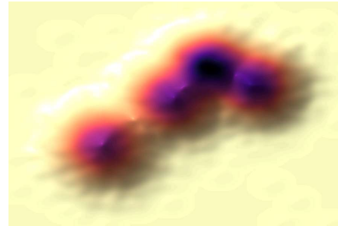
Synthesis2
 # points: 40,040

CDP



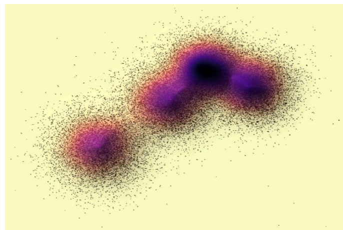
T1: 0.3750 | T2: 0.0574 | T3: 0.7125

IDP



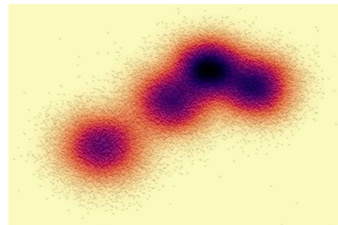
T1: 0.1875 | T2: 0.0987 | T3: 0.5125

SUP



T1: 0.1375 | T2: 0.0864 | T3: 0.0750

VIDP



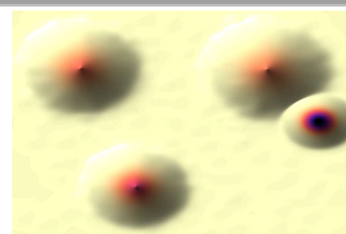
T1: 0.1000 | **T2: 0.0505** | **T3: 0.0625**

Synthesis3
 # points: 160,160

CDP

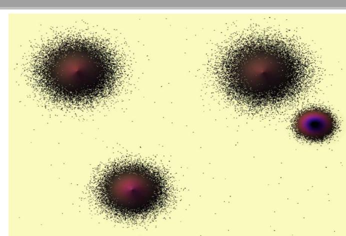
T1: 0.6250 | **T2: 0.0243** | T3: 0.8500

IDP



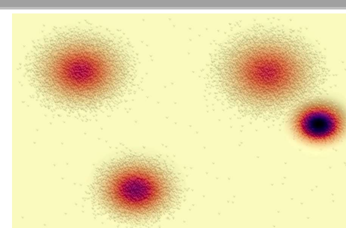
T1: 0.8500 | T2: 0.0356 | T3: 0.6000

SUP



T1: 0.4125 | T2: 0.0294 | T3: 0.0875

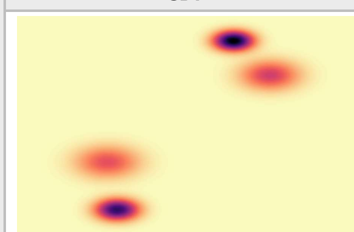
VIDP



T1: 0.3000 | T2: 0.0248 | **T3: 0.0500**

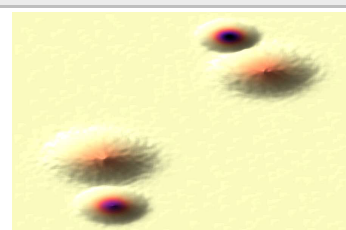
Synthesis4
 # points: 640,640

CDP



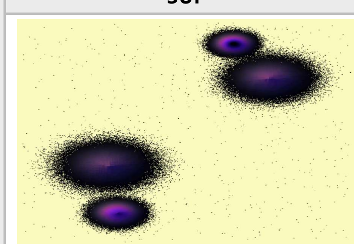
T1: 0.2750 | **T2: 0.0369** | T3: 0.8875

IDP



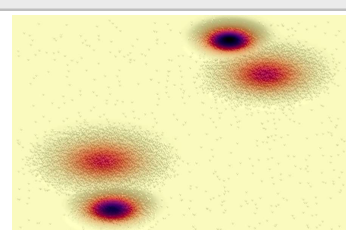
T1: 0.4625 | T2: 0.0935 | T3: 0.7875

SUP

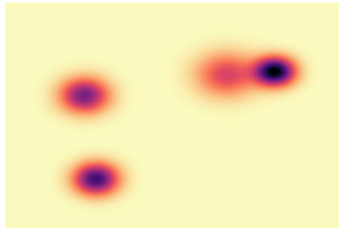
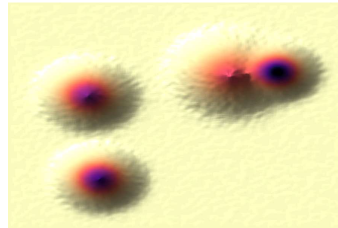
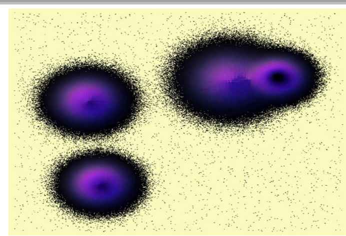
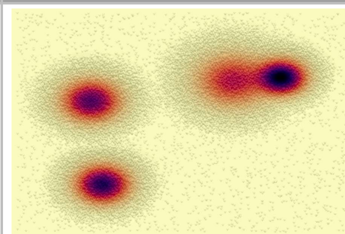


T1: 0.4500 | T2: 0.1766 | **T3: 0.0625**

VIDP



T1: 0.1125 | T2: 0.0654 | T3: 0.0875

<table>
<tr><td rowspan="6">Synthesis5
points:
2,562,560</td><th colspan="3">CDP</th><th colspan="3">IDP</th></tr>
<tr><td colspan="3"></td><td colspan="3"></td></tr>
<tr><td>T1: 0.1000</td><td>T2: 0.0546</td><td>T3: 0.8625</td><td>T1: 0.1875</td><td>T2: 0.0994</td><td>T3: 0.8250</td></tr>
<tr><th colspan="3">SUP</th><th colspan="3">VIDP</th></tr>
<tr><td colspan="3"></td><td colspan="3"></td></tr>
<tr><td>T1: 0.4750</td><td>T2: 0.1372</td><td>T3: 0.0375</td><td>T1: 0.0625</td><td>T2: 0.0618</td><td>T3: 0.0500</td></tr>
</table>

REFERENCES

[1] C. Babusiaux, F. van Leeuwen, M. Barstow, C. Jordi, A. Vallenari, D. Bossini, A. Bressan, T. Cantat-Gaudin, M. Van Leeuwen, A. Brown, et al. Gaia Data Release 2 - Observational Hertzsprung-Russell diagrams. *Astronomy & Astrophysics*, 616:A10, 2018. doi: 10.1051/0004-6361/201832843 2

[2] J. Clore, K. Cios, J. DeShazo, and B. Strack. Diabetes 130-US Hospitals for Years 1999-2008. UCI Machine Learning Repository, 2014. doi: 10.24432/C5230J 3

[3] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi. Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3784–3797, 2017. doi: 10.1109/TNNLS.2017.2736643 3

[4] F. Freitas, F. Barbosa, and S. Peres. Grammatical Facial Expressions. UCI Machine Learning Repository, 2014. doi: 10.24432/C59S3R 4

[5] J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(03):90–95, 2007. doi: 10.1109/MCSE.2007.55 1

[6] Kaggle Inc. Kaggle. `https://www.kaggle.com/`. 1, 6

[7] G. Kindlmann and C. Scheidegger. An Algebraic Process for Visualization Design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2181–2190, 2014. doi: 10.1109/TVCG.2014.2346325 1

[8] A. Mayorga and M. Gleicher. Splatterplots: Overcoming Overdraw in Scatter Plots. *IEEE Transactions on Visualization and Computer Graphics*, 19(9):1526–1538, 2013. doi: 10.1109/TVCG.2013.65 2

[9] A. Srinivasan. Statlog (Landsat Satellite). UCI Machine Learning Repository, 1993. doi: 10.24432/C55887 5

[10] T. Trautner, F. Bolte, S. Stoppel, and S. Bruckner. Sunspot Plots: Model-based Structure Enhancement for Dense Scatter Plots. *Computer Graphics Forum*, 39(3):551–563, 2020. doi: 10.1111/cgf.14001 2

[11] V. Vidulin, M. Lustrek, B. Kaluza, R. Piltaver, and J. Krivec. Localization Data for Person Activity. UCI Machine Learning Repository, 2010. doi: 10.24432/C57G8X 4