# Twitter Thread by Jia-Bin Huang

**Jia-Bin Huang**
@jbhuang0604

**Sharing one idea I found useful for paper writing:**

**Do NOT ask people to solve correspondence problems.**

**Some Dos and Don'ts examples below:**

***Figures*: Don't ask people to match (a), (b), (c) ... with the descriptions in the figure caption.**

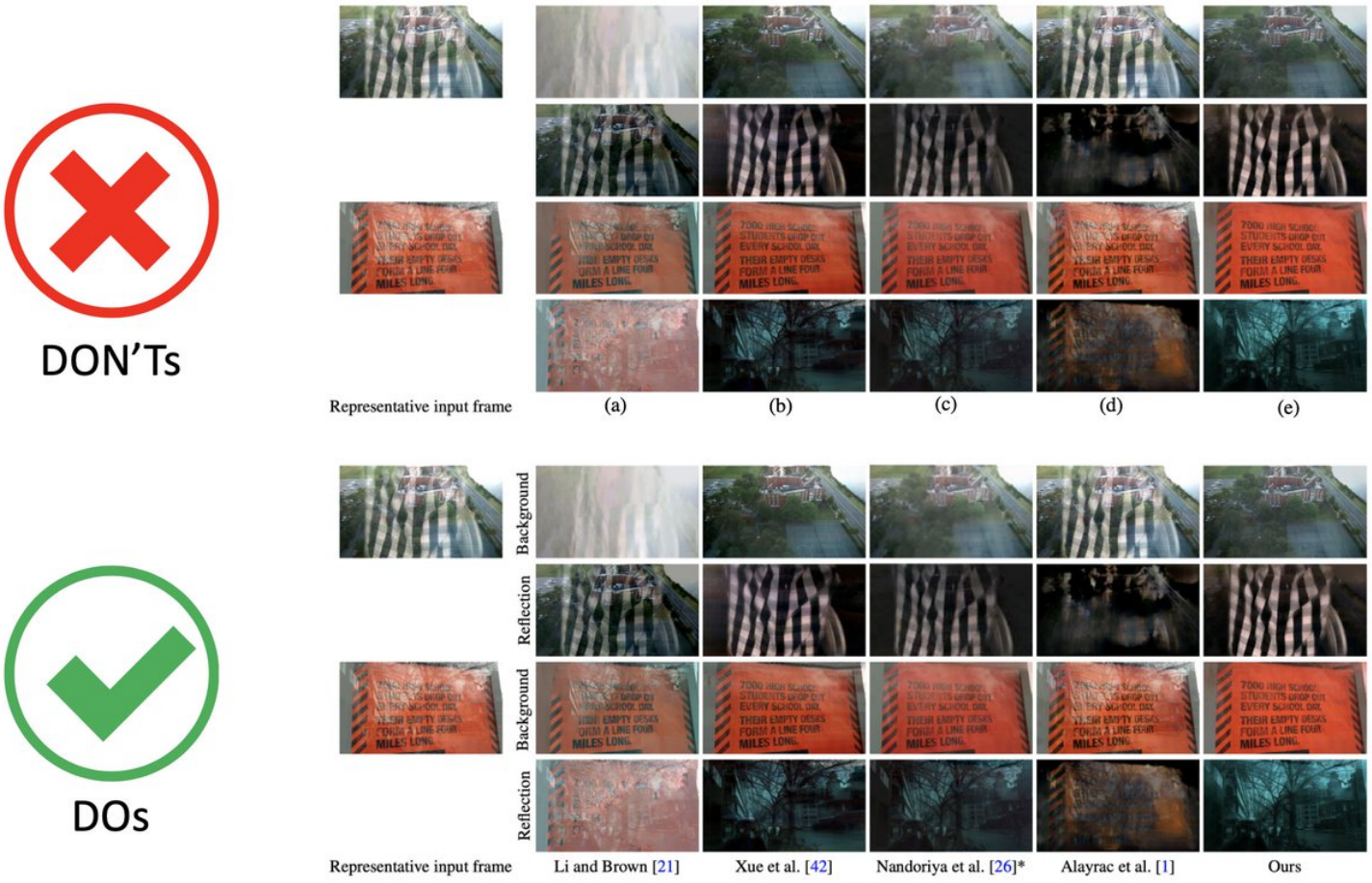



*Figure caption*

Use "self-contained" caption. It's annoying to dig into the texts and match them to the figures. Ain't nobody got time for that! ■■

Also, add a figure "caption title" (in bold fonts). It allows readers to navigate through figures quickly.



DON'Ts

Please refer to the texts for detail.



DOs

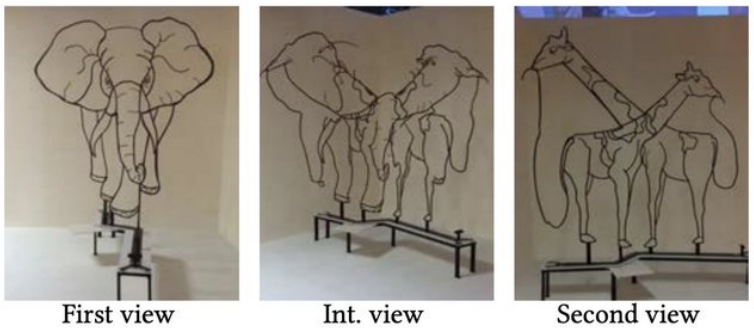



Fig. 2. **Example of multi-view wire sculpture art.** The anamorphose sculpture created by the French sculptor Matthieu Robert-Ortis is a classic example of *multi-view wire art*. When viewing from one specific angle, we perceive a drawing of an elephant. When viewing from another view point, the interpretation changes into two giraffes. The 2D projection in the intermediate view does not produce an interpretable image.

*Notations*

Give specific, meaningful names to your math notations. For example, the readers won't need to go back and forth to figure what each term means.



DON'Ts

$$L = \lambda_1 L_1 + \lambda_2 L_2$$
$$\vdots$$
$$\lambda_1 = 0.1, \lambda_2 = 2.5$$



DOs

$$L_{total} = \lambda_{rec} L_{rec} + \lambda_{smooth} L_{smooth}$$
$$\vdots$$
$$\lambda_{rec} = 0.1, \lambda_{smooth} = 2.5$$

*Which*

I found that many of my students love to use "which" in their sentences. I hate it ... because I often cannot figure out what exactly "which" refers to. Break it down into simple sentences and spell out what that subject of the sentence is.



DON'Ts

We present a method that addresses A, B, C, which is challenging due to X.

? ? ?



DOs

We present a method that addresses A, B, C. The problem A is challenging due to X.

*Respectively*

It's hard to parse which corresponds to which in the sentence that ends with "respectively" (have to solve a long-range correspondence problem). Break them them so that one sentence talks about one thing.



DON'Ts

We show results on three tasks. We improve the performance on the task A, task B, and task C by X%, from Y1 to Y2 mAP, and a Z point gain, respectively.



DOs

We show results on three tasks: On task A, we improve the performance by X%. On task B, we get a boost from Y1 to Y2 mAP. On task C, we achieve a Z point gain.

*Citations*

People like to use many acronyms for their methods. It may be hard for readers to memorize/match which method/dataset/metric you are referring to. Adding citations is an easy way to fix this.



DON'Ts

We compare our method against X-Net, Y-Net, Z-Net on the dataset W.



DOs

We compare our method against X-Net [1], Y-Net [2], Z-Net [3] on the dataset W [4].

*Names for notations*

When using notations in the sentences, mention their "names" as well. The readers won't need to flip through your papers to look up what these notations mean.



DON'Ts

The $F_\theta$ takes $I_i, I_j$ as inputs and produce $W_{i\to j}$.



DOs

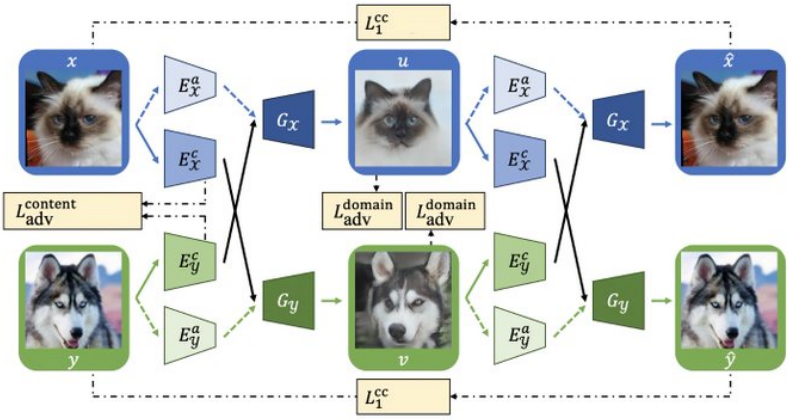
The flow estimation network $F_\theta$ takes a pair of frames $I_i, I_j$ as inputs and produces a dense flow field $W_{i\to j}$.

*Connect figures with equations, notations, and sections*

I view the overview figure in a paper a centralized hub that connects all the important equations, notations, and sections in one place. This makes it easy for people to understand how everything fits together.

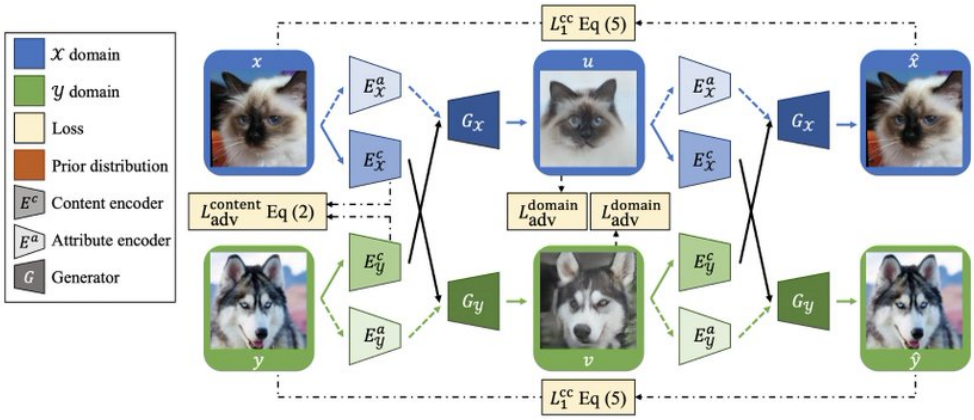


DON'Ts






DOs




*Tables*

Factorize the variants/attributes of different methods so that it becomes clear to compare one with another.



DON'Ts

| Method | mAP |
|---|---|
| [Snorlax et al. 2018]* | 25.0 |
| [Bulbasaur et al. 2019]*† | 29.8 |
| [Psyduck et al. 2020] † | 32.1 |
| Ours | 35.5 |



DOs

| Method | External data? | Finetuned? | mAP |
|---|---|---|---|
| [Snorlax et al. 2018] | ✓ | - | 25.0 |
| [Bulbasaur et al. 2019] | ✓ | ✓ | 29.8 |
| [Psyduck et al. 2020] | - | ✓ | 32.1 |
| Ours | - | - | 35.5 |

*One table, one message*

Decompose your big table so that each table conveys exactly one thing. This avoids people from having to compare results from distant rows. Having multiple smaller tables gets the point across easier. (Don't worry about the redundancy.)



DON'Ts

| Method | Acc (%) |
|---|---|
| Variant A1 | 75 |
| Variant A2 | 65 |
| w/o B | 77 |
| w/o C | 73 |
| Full model | 80 |



DOs

| Method | Acc (%) |
|---|---|
| Full model | 80 |
| Variant A1 | 75 |
| Variant A2 | 65 |

| Method | Acc (%) |
|---|---|
| Full model | 80 |
| Without B | 77 |

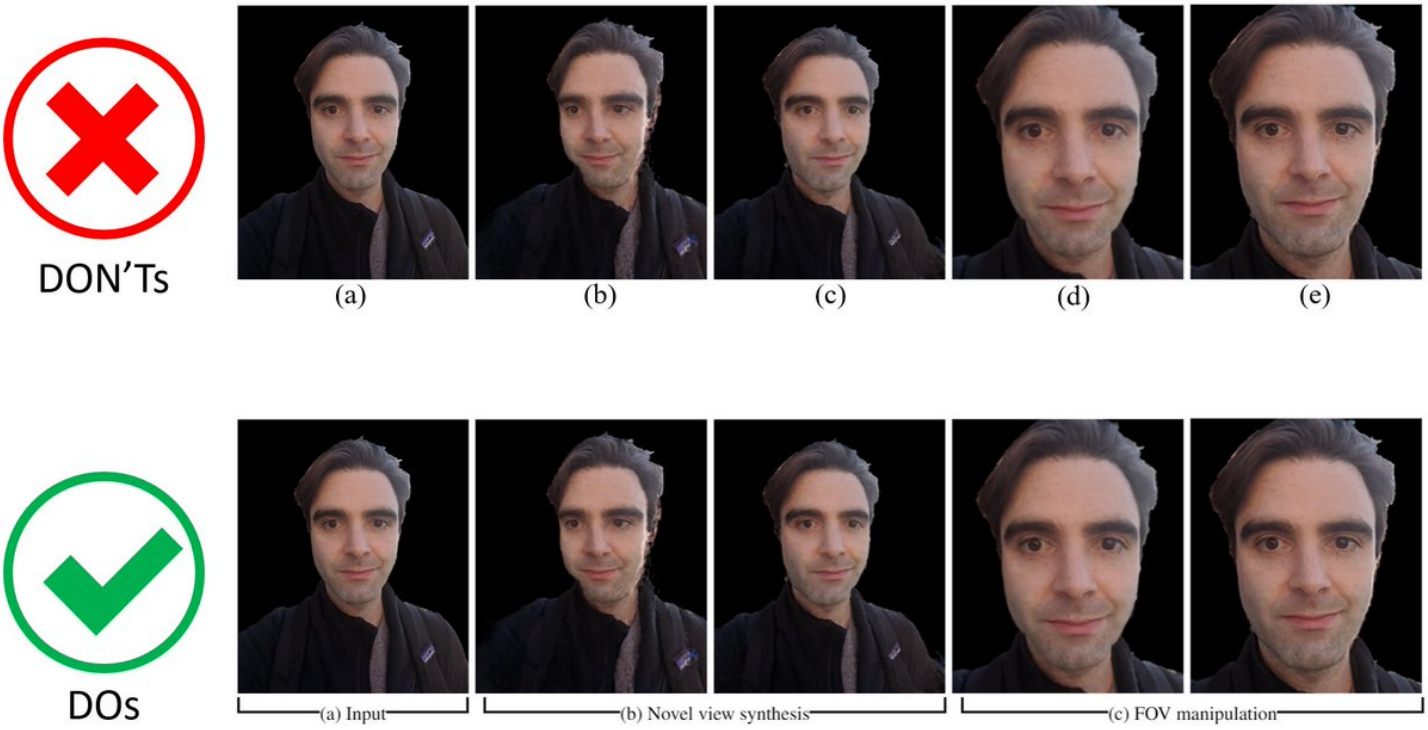
| Method | Acc (%) |
|---|---|
| Full model | 80 |
| Without C | 73 |

*Group subfigures*

Don't ask readers to figure out the grouping (b-c) and (d-e) in the caption when you explicitly group them.

How to create underbracket? Ex:

$\underbracket[1pt][2.0mm]{\hspace{\FIGWIDTH}}_%
{\substack{\vspace{-2.0mm}\\
\colorbox{white}{(a) Input}}}$




*Parallelism*

When applicable, use repetitive grammatical elements in your sentence. It helps the readers to easily parse the parallel concepts you want to convey.



DON'Ts

That's one small step for man,
for mankind it's a one giant leap.



DOs

That's one small step for man,
one giant leap for mankind.

*Table organization*

Merge tables sharing the same structure. Label the metric (the larger/smaller the better) with up-arrow and down-arrow so that your readers don't need to look them up.



DON'Ts

| Method | Abs Rel | Sq Rel | RMSE | log RMSE | $\delta < 1.25$ | $\delta < 1.25^2$ | $\delta < 1.25^3$ |
|---|---|---|---|---|---|---|---|
| DeepV2D [56] | 0.526 | 3.629 | 6.493 | 0.683 | 0.487 | 0.671 | 0.761 |
| Ours - Single-scale pose (aligned MiDaS) | 0.380 | 2.617 | 5.773 | 0.533 | 0.562 | 0.736 | 0.832 |
| Ours - Single-scale pose + depth fine-tuning | 0.472 | 3.444 | 6.340 | 0.635 | 0.534 | 0.694 | 0.790 |
| Ours - Single-scale pose + depth filter | **0.375** | **2.546** | **5.763** | **0.530** | **0.569** | 0.738 | 0.835 |
| Ours - Flexible pose | 0.379 | 2.702 | 5.795 | 0.533 | 0.565 | 0.744 | 0.836 |
| Ours - Flexible pose + depth fine-tuning | 0.439 | 3.100 | 6.213 | 0.614 | 0.524 | 0.698 | 0.796 |
| **Ours - Flexible pose + depth filter** | 0.377 | 2.657 | 5.786 | 0.531 | 0.568 | **0.745** | **0.837** |

| Method | ATE (m) | RPE Trans (m) | RPE Rot (deg) |
|---|---|---|---|
| DeepV2D [56] | 0.9526 | 0.3819 | 0.1869 |
| Ours - Single-scale pose (aligned MiDaS) | 0.1883 | 0.0806 | 0.0262 |
| Ours - Single-scale pose + depth fine-tuning | 0.1686 | 0.0724 | 0.0139 |
| Ours - Single-scale pose + depth filter | 0.1882 | 0.0806 | 0.0262 |
| Ours - Flexible pose | 0.1843 | 0.0723 | 0.0095 |
| Ours - Flexible pose + depth fine-tuning | **0.1656** | **0.0651** | **0.0070** |
| **Ours - Flexible pose + depth filter** | 0.1843 | 0.0723 | 0.0095 |



DOs

| | Depth - Error metric↓ | | | | Depth - Accuracy metric↑ | | | Pose - Error metric↓ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | Abs Rel | Sq Rel | RMSE | log RMSE | $\delta < 1.25$ | $\delta < 1.25^2$ | $\delta < 1.25^3$ | ATE (m)↓ | RPE Trans (m)↓ | RPE Rot (deg)↓ |
| DeepV2D [56] | 0.526 | 3.629 | 6.493 | 0.683 | 0.487 | 0.671 | 0.761 | 0.9526 | 0.3819 | 0.1869 |
| Ours - Single-scale pose (aligned MiDaS) | 0.380 | 2.617 | 5.773 | 0.533 | 0.562 | 0.736 | 0.832 | 0.1883 | 0.0806 | 0.0262 |
| Ours - Single-scale pose + depth fine-tuning | 0.472 | 3.444 | 6.340 | 0.635 | 0.534 | 0.694 | 0.790 | 0.1686 | 0.0724 | 0.0139 |
| Ours - Single-scale pose + depth filter | **0.375** | **2.546** | **5.763** | **0.530** | **0.569** | 0.738 | 0.835 | 0.1882 | 0.0806 | 0.0262 |
| Ours - Flexible pose | 0.379 | 2.702 | 5.795 | 0.533 | 0.565 | 0.744 | 0.836 | 0.1843 | 0.0723 | 0.0095 |
| Ours - Flexible pose + depth fine-tuning | 0.439 | 3.100 | 6.213 | 0.614 | 0.524 | 0.698 | 0.796 | **0.1656** | **0.0651** | **0.0070** |
| **Ours - Flexible pose + depth filter** | 0.377 | 2.657 | 5.786 | 0.531 | 0.568 | **0.745** | **0.837** | 0.1843 | 0.0723 | 0.0095 |

*Shape attributes*

Leverage the shape attributes (color, thickness) to encode the message.

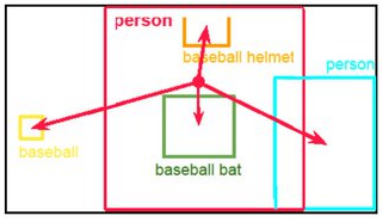
Also, use a deemphasized image in the background to avoid mental matching.



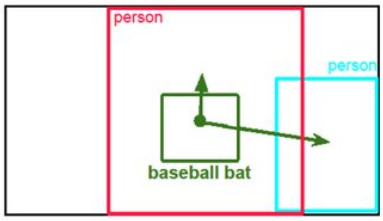
DON'Ts




(a) Object detection




(b) Human-centric



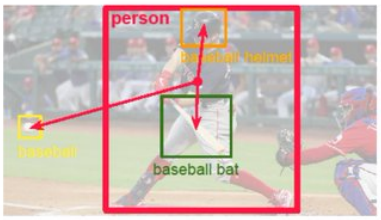

(c) Object-centric



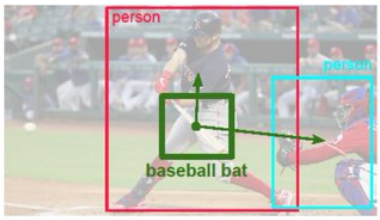
DOs




(a) Object detection




(b) Human-centric




(c) Human-centric