Visual-Aware Testing and Debugging for Web Performance Optimization

*Xinlei Yang, Wei Liu, **Hao Lin**, Zhenhua Li,*

Feng Qian, Xianlong Wang, Yunhao Liu, Tianyin Xu

清華大學 Tsinghua University

UNIVERSITY OF MINNESOTA

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Web performance optimizers (WPOs)

- Optimizing web page loading by image transcoding, JavaScript/CSS minification, HTML compression, etc.
- Saving the page load time (by 2.5x) and network traffic (by 2-3x)
- Used on **5 billion** web pages, benefiting **tens of millions** of users

Google AMP

Baidu TrafficGuard

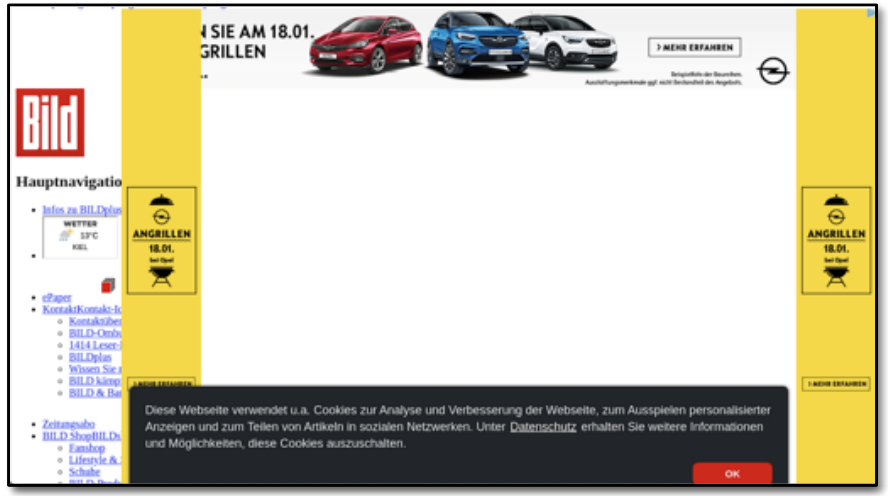
Google Flywheel

Ziproxy

How reliable are WPOs in practice?

Original web page

Ziproxy



"Optimized" version

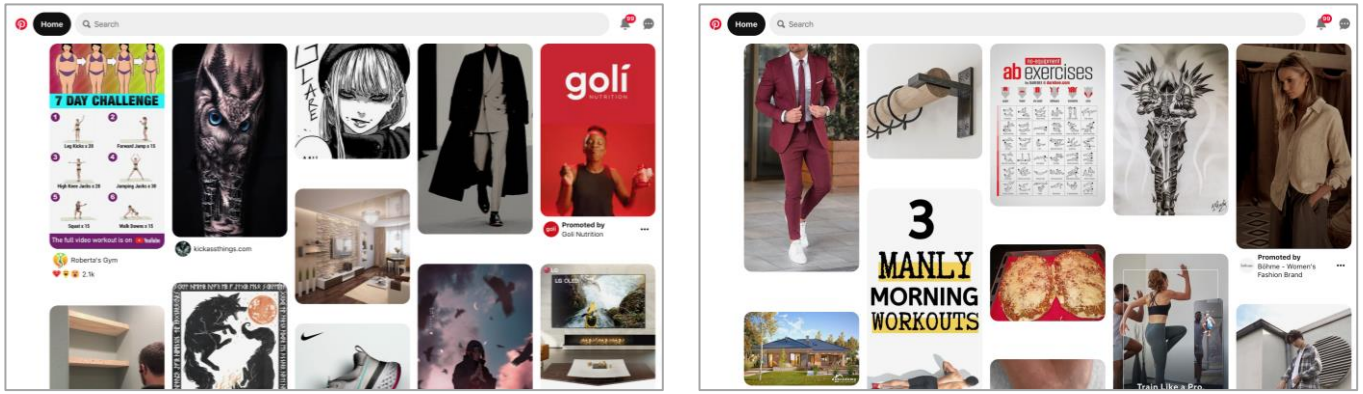
WPOs can induce visual distortions on web pages!

Visual distortions are hard to detect and debug

- **Detection**: **dynamic contents** vary significantly among different loads

Intuitive Method	Problem
CV-based image similarity measurement	Easily affected by dynamic pages
Comparisons of key data structures (DOM & CSSOM)	Lack visual hints, being over-general

- **Debugging**: **sophisticated implementations** of WPOs & **limited information** for pinpointing root causes



Dynamic contents among different loads

~20K LoCs

40+ third-party modules

Very few runtime logs

Implementation of Ziproxy

Contributions

- The first user study on visual distortions incurred by WPOs and **dataset release** involving 5,000 websites

- Vetter**:** visual-aware testing and debugging for WPOs
  - **Key idea**: exploiting the visual morphology of web pages
  - **Open sourced** at https://github.com/Web-Distortion/Vetter

- Detected 21 unknown defects in four widely used WPOs
  - 13 are confirmed and 6 are fixed

Understanding visual distortions

- Crowdsourcing study on Ziproxy and Compy involving **5K** pages and **18** users
- Ziproxy and Compy incur visual distortions on **3.3%** and **6.1%** pages
- For **93%** web pages, the inspectors have the same opinions

Distortion Symptom	# by Ziproxy	# by Compy	Total
Content Loss	0	63	63
Image Display Error	11	0	11
Text Confusion	13	16	29
Layout Disorder	0	3	3
All	**24**	**82**	**106**

Visual distortions occurring to the Alexa top 2,500 websites' landing pages

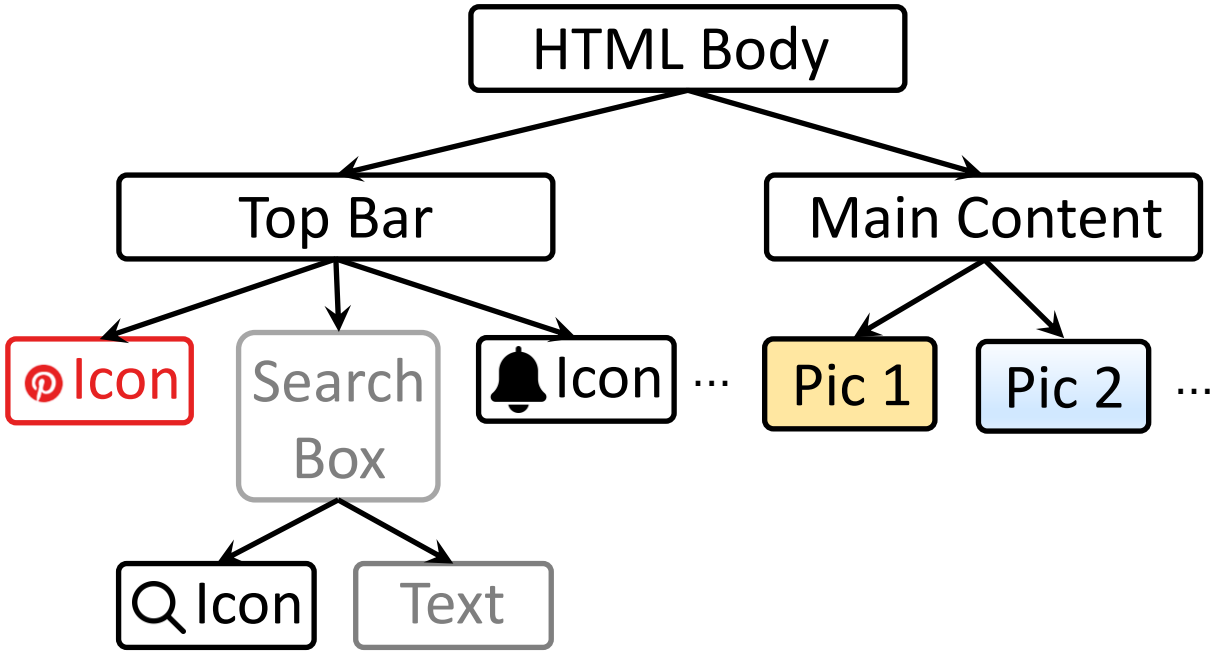
Diagnosing visual distortions with a page's morphology

- Web frameworks mostly follow the Model-View-ViewModel (MVVM) pattern
- While the model (logic and data) is dynamic among loads, visual elements' **topological forms** & **scale-free geometrical structures** are stable

Render result

```
<body>
    <div id="topBar">
        <div id="barLayout">
            <div class="transparentContainer">
                ...
                    <div id="logo">
                        <a href="logo.png"> ... </a>
                    </div>
                    <div id="searchBoxContainer">
                        <svg href="search.png"> ... </svg>
                        <input placeholder="Search"> ... </input>
                    </div>
                    <div>
                        <a href="bell.png"> ... </a>
                    </div>
                ...
            </div>
        </div>
        <div id="mainContent">
            ...
                <div id="pic_1">
                    <img src="https://...">
                </div>
                <div id="pic_2">
                    <img src="https://...">
                </div>
            ...
        </div>
    </div>
</body>
```

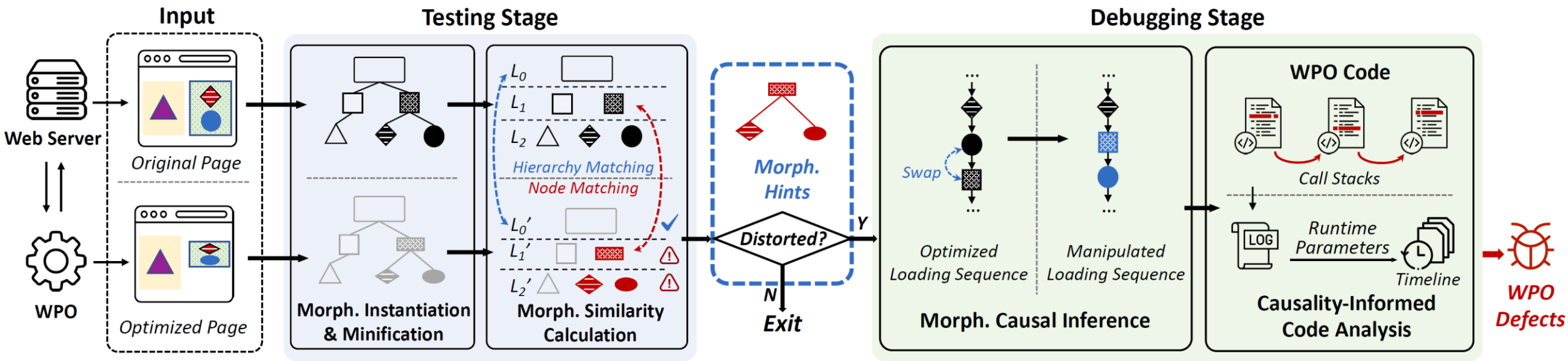
Source code



Web page's morphology

Vetter: automatic testing & debugging for WPOs

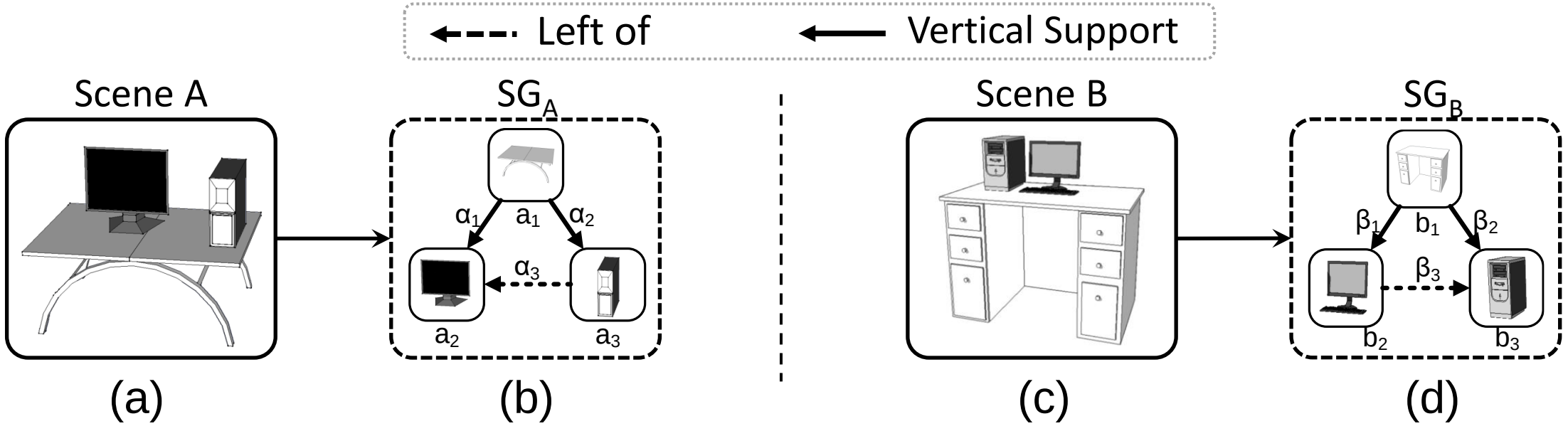
- **Testing**: measure the similarity of web pages' morphologies
- **Debugging**: localize offending elements & WPO bugs via **"morph. hints"**



Architectural overview of Vetter

Testing: extracting morphology with scene graph

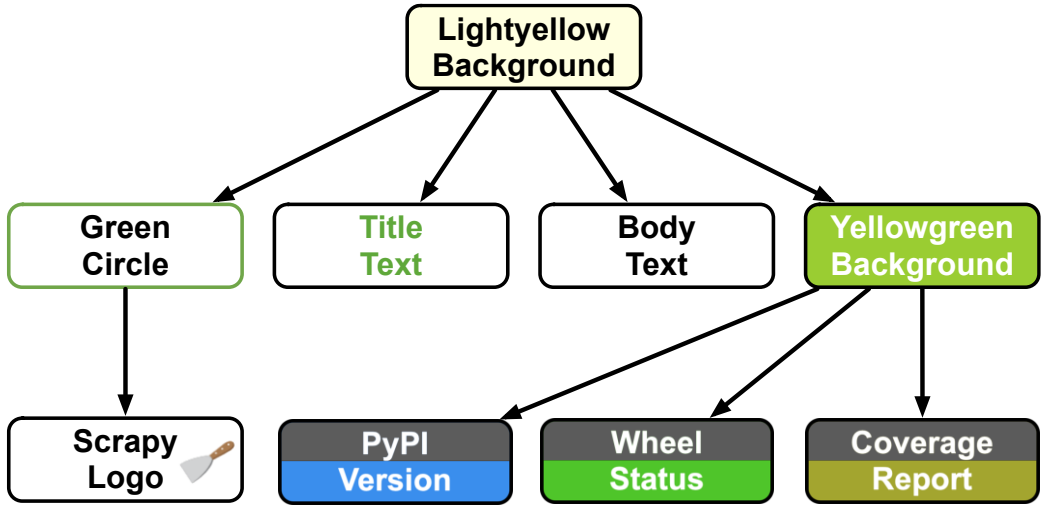
- Represent web pages' morphologies using **scene graph**
- A classic data structure in computer graphics for representing 2D/3D scenes



Testing: reducing the complexity of the scene graph

- Minimize scene graph into a **morphological segment tree (MST)**
- Based on **intrinsic hierarchy** in a web page's morphology

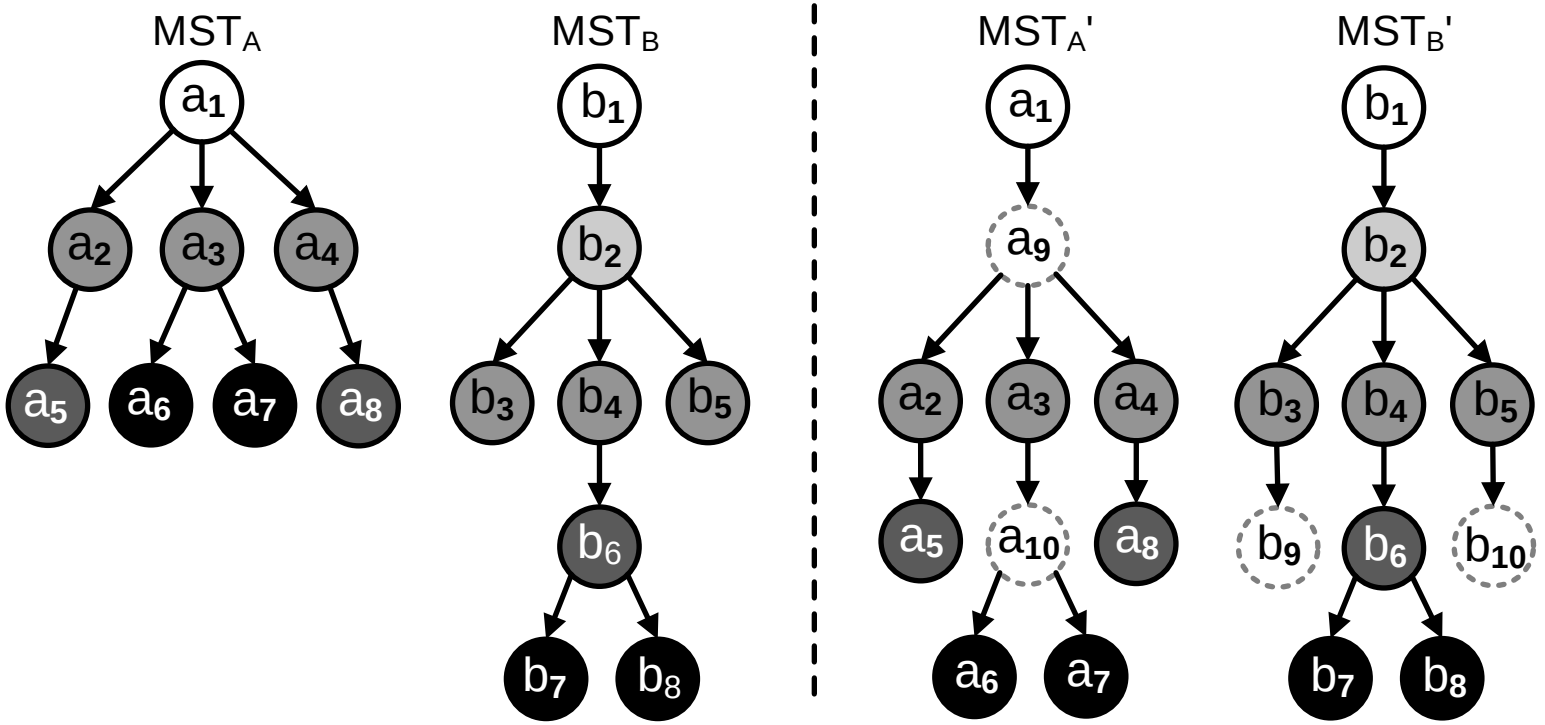
(a) Rendered Page



(b) MST

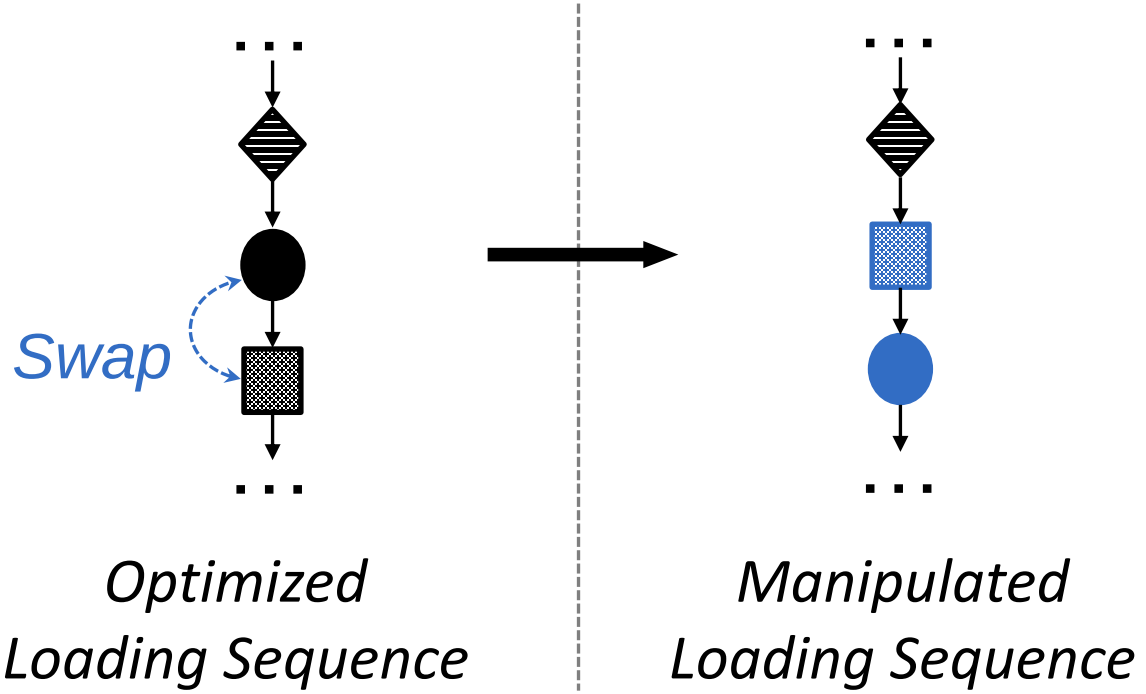
Testing: calculating the similarity of morphologies

- Match MSTs of the web page before and after a WPO
- Coarse-grained level-by-level matching
- Fine-grained node-by-node matching



Debugging: morphological causal inference

- Many visually distorted elements are the “chain reaction” results of neighboring visually distorted elements
- **Restore** the optimized web page resources/loading sequences
- **Test** again to check whether the distortion is resolved



Evaluation and real-world impacts

- Discovered **21** previously unknown defects of **4** representative WPOs
- 13 are confirmed and 6 are fixed

Compy

ID	Description	Issue/PR NO.	Current State
1	Fail to compress JPG/PNG images.	Issue-63 & PR-70	**Confirmed & Fixed**
2	Fail to parse compressed images.	Issue-64	Reported
3	Can't deal with websocket.	Issue-65	Reported
4	Block redirecting process of websites.	Issue-66 & PR-68	**Confirmed & Fixed**
5	Can't support GIF images.	PR-70	**Confirmed & Fixed**

Ziproxy

ID	Description	Issue/PR NO.	Current State
1	Fail to compress some contexts (i.e., generating messy code).	-	Reported
2	Disturb loading sequence of JS files.	-	Reported
3	Cannot handle GIF files.	-	Reported
4	Cause conflicting fields in response header.	-	Reported

Fawkes

ID	Description	Issue/PR NO.	Current State
1	Can't handle elements whose innerText has multiple lines.	Issue-14	Reported
2	Fail to select elements in template HTML.	Issue-13	Reported

SipLoader

ID	Description	Issue/PR NO.	Current State
1	Can't track dependencies triggered by CSS files.	Issue-1	**Confirmed**
2	Can't handle dependency loops among resources.	Issue-2	**Confirmed**
3	Can't request cross-origin resources.	Issue-3	**Confirmed**
4	Disordered page loading of websites with multiple HTML files.	Issue-4	**Confirmed**
5	"404 Not Found" error when loading websites with multiple HTML files.	Issue-5	**Confirmed**
6	Can't handle some dynamic resources.	Issue-6	**Confirmed**
7	Issue related to Chromium.	Issue-7	**Confirmed**
8	CSS abormality of some websites.	Issue-8 & PR-9	**Confirmed & Fixed**
9	Can't rewrite Brotli-compressed contents.	Issue-10 & PR-12	**Confirmed & Fixed**
10	Can't distinguish between data URIs and real URLs in CSS files.	Issue-11 & PR-9	**Confirmed & Fixed**

Conclusion

- A large study on visual distortions from users' perspective
- A novel testing and debugging method for WPOs' incurred visual distortions based on visual morphology
- Detect and fix real-world bugs of representative WPOs
- Open source dataset and code at https://github.com/Web-Distortion/Vetter

link to artifacts