A Comparative Evaluation of Next-Generation Image Formats on Low-Cost Mobile Hardware

Shaurya Singh
Computer Science, NYU Abu Dhabi
shaurya.singh@nyu.edu

Advised by: Yasir Zaki

ABSTRACT

This study presents a thorough evaluation of next-generation WebP and AVIF image formats in comparison to the established JPEG and PNG standards, specifically targeting resource-constrained, low-cost hardware. We seek to quantify the performance improvements resulting from the transition to these advanced formats across an extensive range of web pages and investigate the variations in gains based on numerous factors, such as page size, economic context of the country of origin, and web page popularity. The paper utilizes a set of 750 web pages originating from 30 countries, equally split between low-income and high-income countries that together account for 75% of the world's population. Furthermore, a secondary analysis is run on a statistically representative set of 1200 web pages selected from the world's top 1 million websites to help measure the performance gains offered by the newer formats as a function of the page's popularity.

Reference Format:
Shaurya Singh. 2023. A Comparative Evaluation of Next-Generation Image Formats on Low-Cost Mobile Hardware. In *NYUAD Capstone Project 2 Reports, Spring 2023, Abu Dhabi, UAE.* 8 pages.

1 INTRODUCTION

Google's Next Billion Users research project estimates that within the next four years, a billion more people will purchase their first smartphone [1]. This trend is likely to continue as mobile data rates also continue to decrease [2], making it more affordable for more people to access the internet through their mobile devices. In order to meet the needs of this growing user base, it will be necessary to invest in increasing the Internet's load-carrying capacity through major capital investments in the long term. However, in the short term, one potential solution is to focus on creating more efficient image assets that are more suited for low-bandwidth networks servicing devices with limited hardware resources.

An analysis of the HTTP Archive's set of over 1 million web pages using BigQuery suggests that 82% of web-pages use at least one PNG image, whereas a whopping 91% of web-pages in the set use at least one JPEG image. There is no doubt that images are virtually omnipresent on the internet – which also makes them a perfectly suited target for further optimization techniques in order to reduce network loads and page speeds. [3]

To further get an accurate picture of our problem, we delve deeper into analyzing web pages. The median mobile page weight has increased from 295KB in January 2012 to 1993KB in April 2023, an increase of 575% in slightly over a decade. The 75th percentile mobile page weight stands at 4062KB, and this number rises to 8195KB for the 90th percentile. The median mobile web page contains 677KB of image data, which accounts for over a third of all page weight.

This project aims to explore two new next-generation image formats: WebP and AVIF, that promise significant performance and bandwidth gains over the incumbent PNG/JPEG formats. By measuring the impact on multiple web page metrics across a range of websites, we seek to evaluate the feasibility and performance of these image formats in order to find patterns and ideal use-cases where they deliver maximum impact.

1.1 Next-Generation Image Formats

The WebP (*Web Picture*) image format, developed by Google, represents an open, royalty-free solution designed to replace

This report is submitted to NYUAD's capstone repository in fulfillment of NYUAD's Computer Science major graduation requirements.

Capstone Project 2, Spring 2023, Abu Dhabi, UAE

existing image formats like JPEG, PNG, and GIF. This progressive image file format offers superior image compression while preserving high-quality visuals, all within small file sizes, which makes it an ideal choice for web use.

The AVIF (*AV1 Image File Format*) is an new image file format that utilizes the advanced AV1 video codec to compress and encode image data. This cutting-edge format was developed by the Alliance for Open Media (AOMedia), with the vision of providing an open, royalty-free alternative to existing, often restrictive, image formats such as JPEG, PNG, and GIF.

2 RELATED WORK

A significant amount of academic work exists on the both formats' performance in terms of image quality as opposed to legacy standards such as JPEG or PNG. Previous research on AVIF's size advantage over legacy formats *while maintaining quality* has made significant conclusions, with one study finding that AVIF "results in the best overall performance [and bitrate savings] considering both 4:2:0 and 4:4:4 chroma sub-sampling encoded images" [4] and also outperformed WebP in the cited study. Outside of academia, the technology industry has been quite optimistic about AVIF as well. In 2020, Netflix began moving most of their image assets to an AVIF-first approach, with JPEG and PNG files as fallback options. Netflix's large scale experimentation with AVIF concluded "superior compression efficiency" compared to JPEG, with similar picture quality in the end product. [5].

WebP on its own is a major candidate for the web's foremost image format, as it has a significantly better data compression algorithm compared to JPEG and PNG, for only a slightly worse compression quality [7]. Previous research has also found that WebP shows notable efficiency when it comes to compressing natural images, often slightly outperforming other image codecs in terms of both Peak Signal-to-Noise Ratio (PSNR) and Human Visual System (HVS)-based evaluations. This could make it an effective choice for applications where natural images need to be compressed to lower bitrates. The current implementation of WebP is not particularly efficient in the coding phase, although it performs reasonably well in decoding low-resolution images. This could impact its usefulness in scenarios where speedy encoding is a crucial requirement. [6]

3 METHODOLOGY

To run a full-fledged analysis of how the WebP and AVIF formats compare especially in the context of mobile web pages running on low-end hardware, we need to measure their performance on three fronts:

- **Comparative Performance Gains by Income Classification**, which refers to comparing how the multiple measures of web page performance fare for websites from a basket of low and high income economies. We measure how identical web pages containing images from each format perform on a network in terms of Speed Index, the Page Size, Page Load Time, First Contentful Paint, as well as the Document Complete Time. A small section detailing these metrics is included later in the report.
- **Representative Set Study**, where we cover a small set that statistically mirrors the properties of the set of the top 1 million web pages in the world. This analysis is country and income agnostic, and all we really want to learn is how gains from WebP and AVIF distribute themselves in the context of popularity measured via domain ranking.
- **Image Asset Size Reduction**, which is fairly self-evident. Smaller image sizes generally correlate with faster download and decode speeds. Measuring the output sizes of the image assets created by each technique as opposed to the standard PNG and JPEG formats will give us a good understanding of this metric.

3.1 Preparing the Data-set

As a starting reference point, we use the Tranco List, a list of the web's top 1 million websites that aggregates data over multiple similar lists in order to create a research-oriented top sites ranking that is hardened against manipulation. The Tranco list is more temporally stable and uses data aggregated from the Alexa, Cisco Umbrella, and Majestic lists.

We extract a custom Tranco list with the following properties:

- Each domain included in the custom list should be a part of at least 2 out of 3 base lists.
- We exclude subdomains from our analysis.
- For organizations that have many local TLDs (such as Google), we only consider the single most popular URL and discard the rest.
- We also restrict our domains to those that appear in the Chrome User Experience Report of 2023, which contains URLs popular with Chrome users and therefore reflects browser traffic across a large user base.

3.2 Sampling Countries to Analyze

For this project, we need to sample a diverse set of countries that vary from each other significantly in terms of economic benchmarks. We use the World Bank's Gross National Income (in US$, calculated using the Atlas method) metric, which has been used to classify countries into four categories by income. We pick the world's top 30 most populous countries, which combined account for over 85% of the world's population. Another extremely neat observation is that when the populations of these countries are grouped into two buckets of (Low-Income + Low-Middle Income) and (Upper-Middle Income + High-Income), the distribution of population is fairly even with roughly 2.9 Billion people on both sides of our economic categorization.

The final list of countries is as follows:

- **Low-Income Economies**: Myanmar, Vietnam. (2)
- **Low-Middle Income Economies**: Philippines, Indonesia, Egypt, Iran, Pakistan, Bangladesh, India, Tanzania, Kenya,Nigeria, Democratic Republic of the Congo, Thailand, and Ethiopia. (13)
- **Upper-Middle Income Economies**: China, Turkey, Russia, Colombia, Brazil, Mexico, and South Africa. (7)
- **High-Income Economies**: South Korea, Japan, Spain, Italy, France, United Kingdom, Germany, and the United States. (8)

Table 1: Describing the Representative Set

r_length	s_rank	e_rank	n_picked	% picked
100	1	100	100	100
200	101	300	100	50
400	301	700	100	25
800	701	1500	100	12.5
1600	1501	3100	100	6.25
3200	3101	6300	100	3.13
6400	6301	12700	100	1.56
12800	12701	25500	100	0.78
25600	25501	51100	100	0.39
51200	51101	102300	100	0.2
102400	102301	204700	100	0.1
204800	204701	366410	100	0.05
		Total	1200	

This table provides an overview overview of the representative set including range length (r_length), start rank (s_rank), end rank (e_rank), number picked (n_picked, always equal to 100), and the percentage picked from each range (%picked).

3.3 Representative Set Analysis

We also want to measure the potential impact of the new image formats on the larger set of all the URLs we retrieved from our custom Tranco list. Therefore, we pick a representative smaller set that approximates the statistical characteristics of our original set of websites that we retrieved from Tranco.

This results in a representative dataset that can be described as per the above Table 1.

3.4 Image Level Analysis

In addition to measure metrics pertaining to web page performance, we also look at the set of images that occur in these web pages independently. By converting all the PNG and JPG image issets to WebP and AVIF, we can study the following properties:

- The overall size reductions offered by the next generation image formats over the incumbent standards.
- Trends surrounding these size reductions, and how percentage savings correlate with image asset size to pinpoint if WebP and AVIF are more or less effective with certain image asset sizes than others.

3.5 Technical Details

Here is a step-by-step breakdown of the tools we use in order to cache/clone the target web pages, convert our image assets, and then measure the various performance metrics associated with web pages.

- To cache websites, we use a python-based scraping script that stores all the assets, image or otherwise, on to our local server with a hash table for all assets constructed using an SQL database.
- To convert PNG and JPG image assets to WebP and AVIF, we use the Pillow library, a free and open-source additional library for the Python programming language that adds support for opening, manipulating, and saving many different image file formats.
- To serve original and cloned web pages, we use a man-in-the-middle proxy that intercepts the URL requests and delivers either the "original", or the cloned image versions using WebP/AVIF of the requested web page on our mobile device, which is a Google Pixel 2.
- To measure page metrics, we use WebPageTest, a web page performance testing tool developed by AOL and open-sourced in 2008 and currently maintained by Google on Github.

4 RESULTS

4.1 Income Group Analysis

In this analysis, we first see the macro effects of using WebP and AVIF formats instead of the PNG/JPG formats on the following metrics:

- Page Size, which is self-evident and refers to the total size of the loaded page in Kilobytes.
- Page Load Time, which is the total time it takes for the page to fully load in the browser.

- First Contentful Paint, or FCP, which measures the time from when the page starts loading to when any part of the page's content is rendered on the screen. For this metric, "content" refers to text, images (including background images), <svg> elements, or non-white <canvas> elements.
- Speed Index, which is a page load performance metric measuring how quickly the contents of a page are visibly populated. Speed Index is dependent on size of the viewport and expressed in milliseconds: the lower amount of time the better the score.
- Document Complete Time, which is the time until the document is complete and the browser onLoad event is triggered.

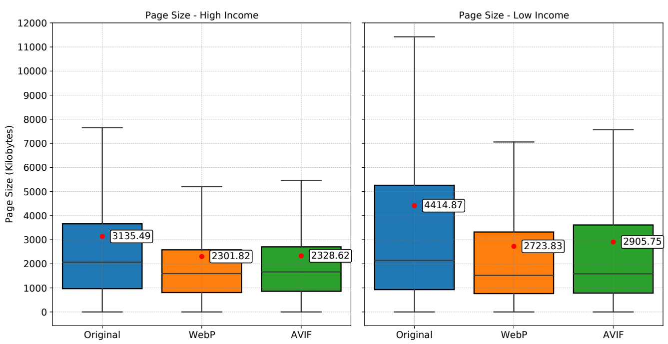

Figure 1: Page Size variation with WebP and AVIF formats, categorized by income group.

4.1.1 *Page Size.*

Figure 1 shows us the Page Size for all our web pages in their original state, as well as the WebP and AVIF cloned versions. We notice that both WebP and AVIF drive significant size gains for both income classifications, with WebP pages performing marginally better. Notice that while the mean values drop by 25% for High Income countries and 35% for Low income countries, the median values do not exhibit a similarly large improvement. This suggests that larger pages stand much more to gain from the switch to WebP and AVIF in terms of image size.

We can also plot a Cumulative Distribution Function table that describes the probability that a randomly selected web page experiences gains greater than or equal to X. Comparing the ratio of the image bytes requested by a web page and the total size of the web page is another interesting metric, as we know that image contribution to the ever-growing page sizes is quite immense. Both these charts can be seen in Figure 3.

Figure 3 (Inset) also shows that the percentage of image bytes as a function of overall page size goes down from 38% on average to about 30%, a decrease that is completely brought out due to the size reductions facilitated by WebP and AVIF.

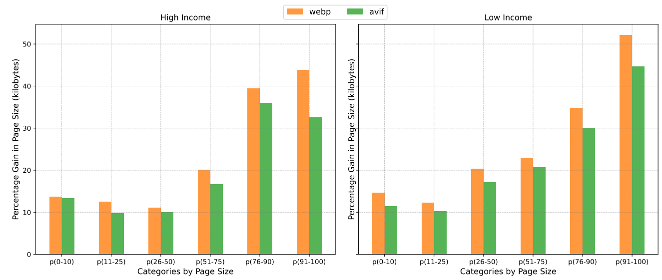

Figure 2: Percentage Gains in page size offered by WebP and AVIF for varying page sizes. Notice how both formats offer increased efficiency at larger pagesizes, and low income countries' web pages typically experience better size reductions. WebP beats AVIF across the board, but both offer significant page size efficiencies.
See table 2 for percentile to size conversion.

Table 2

Percentile	Size Range (KB)
0 - 10	0 - 381
11 - 25	382 - 913
26 - 50	914 - 1993
51 - 75	1994 - 4062
76 - 90	4063 - 8195
91 - 100	> 8195

Converting percentiles to Size Ranges.

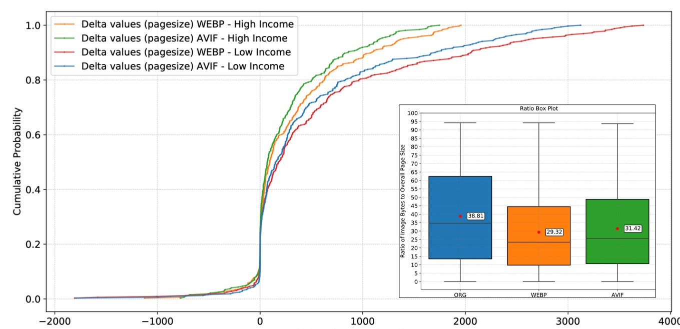

Figure 3: A Cumulative Distribution Function that describes the probability of a randomly selected web page experiences gains greater than or equal to X, with probability on the y-axis and X on the x-axis.
Inset: A box plot visualizing the ratio of image size and total page size across the data set when original (JPG/PNG), WebP and AVIF images are used.

Table 3

Speed Index Range	Interpretation
0 - 3.4	Fast
3.4 - 5.8	Moderate
Over 5.8	Slow

Interpreting Speed Index values

Table 4: Web Pages from High Income Countries

Type	Slow (%)	Moderate (%)	Fast (%)
Original (JPG/PNG)	62	29	9
WebP	48	30	22
AVIF	47	31	22

Variation in Speed Index values

Table 5: Web Pages from Low Income Countries

Type	Slow (%)	Moderate (%)	Fast (%)
Original (JPG/PNG)	58	31	11
WebP	43	29	28
AVIF	43	33	24

Variation in Speed Index values

4.1.2 Speed Index. The Speed Index, a critical measure of web performance, quantifies the time required for the visible parts of a web page to load. Using the interpretation in Table 3 (courtesy Google Chrome Developers Resources), we categorize our web pages from High Income and Low Income countries and measure the percentage distribution for each categories. Then, we can construct the same chart by swapping out the original JPG/PNG images for WebP and AVIF images to see how the percentage distribution varies.

Table 4 and 5 show the variations in Speed Index classification when original images (JPG/PNG) are replaced by WebP and AVIF in web pages originating in High Income countries. These statistics paint a significantly promising picture of the impact WebP and AVIF can have on a web page's Speed Index.

In both cases, the number of pages marked as Moderate stays relatively unchanged. Drawing a CDF which plots the distribution of the absolute gains made as a result of the WebP and AVIF formats also gives us encouraging results, as seen in Figure 6.

4.1.3 First Contentful Paint. Figures 7 and 8 show the impact of WebP and AVIF on the FCP values of both sets of web

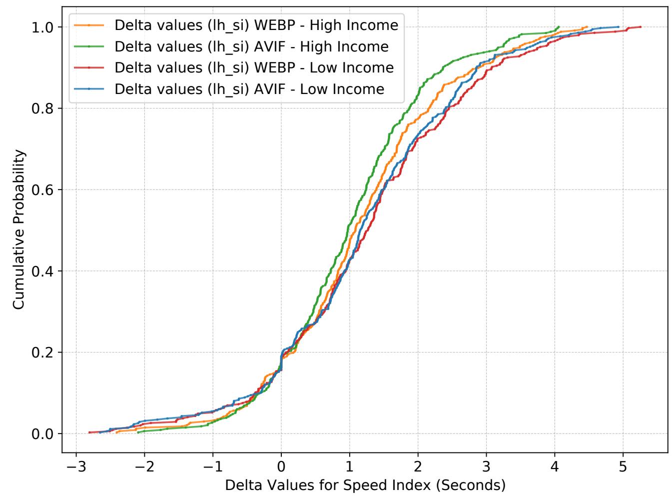

Figure 4: CDF of the gains in Speed Index. 80% of the webpages experience a reduction in Speed Index for both formats and both income groups. 60% will have the SI reduce by 1 second or greater.

pages. Once again, we see encouraging results for both sets, and slightly better results for the low income group.

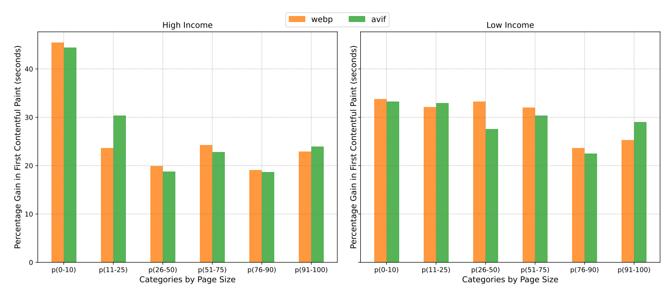

Figure 5: Percentage Gains in page size offered by WebP and AVIF for varying page sizes. Lower Income group typically sees better results, especially for medium to large sized pages. Nevertheless, both categories see significant improvements to FCP.

4.1.4 Page Load Time and Document Complete Time. In this section we take a look at the Page Load Time and Document Complete Time metrics. While the previous metrics have shown encouraging results, these measurements reveal just how efficient the legacy formats are at encoding and decoding, meaning improvements are extremely muted for large pages across income groups. However, as shown in Figure 9, >80% of web pages still experience a net reduction in page load time, with 60% of the pages with a 15% or higher reduction. These are promising numbers, but are not reproduced at the tail end of the data set in large pages.

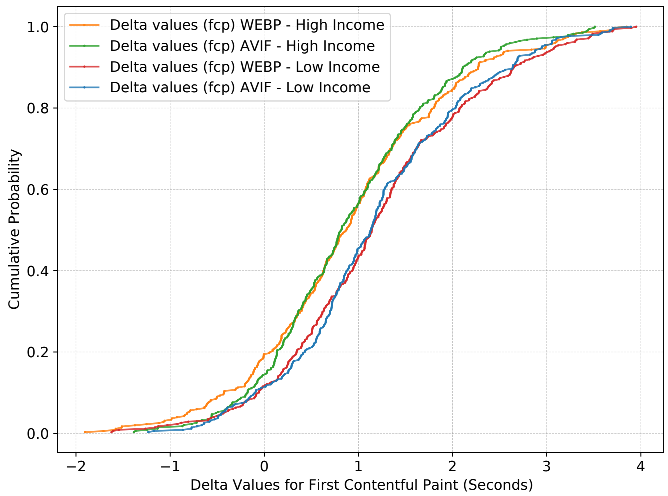

Figure 6: CDF plotting FCP delta values, confirming that Lower Income group gets better gains in FCP.

As shown in Figure 10, Document Complete Time follows a similar pattern to Page Load Time, with 80% of the web pages seeing an improvement and roughly 60% seeing an improvement of 20% or higher.

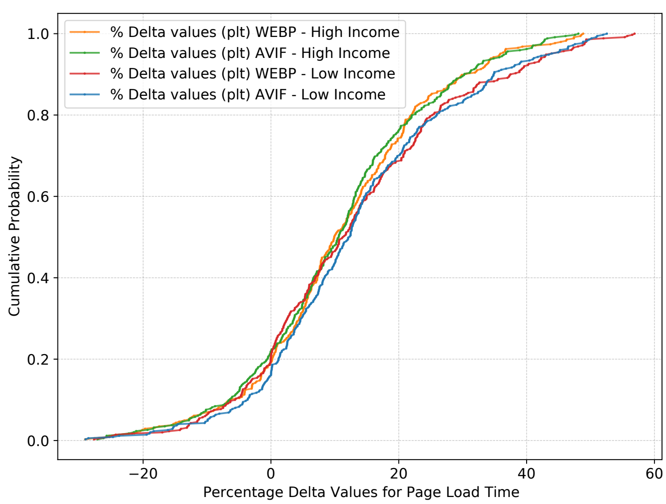

Figure 7: Page load time sees improvements for > 80% of webpages across Income groups, with 60% pages seeing an improvement of 15% or above, but this effect is extremely muted in the largest 10% of pages, which see a negative effect for High Income group and breakeven for Low Income (See Appendix).

4.2 Representative Set Analysis

In this analysis, we see a familiar pattern emerging, as the top 100 ranked websites generally suffer the most when changed

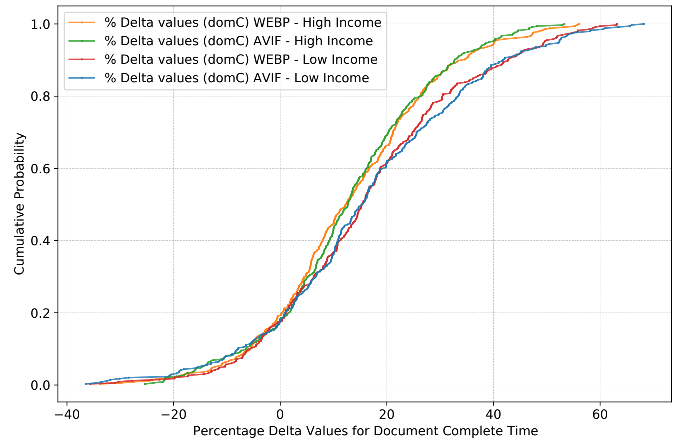

Figure 8: CDF for Percentage Improvements in Document Complete Time

over to WebP or AVIF. This may be due to their high levels of prior optimization, meaning any tampering with the page results in worse outcomes.

Figures 11, 12 and 13 respectively chart the percentage improvements offered by WebP and AVIF for page size, Speed Index and Document Complete Time, where we see the trend emerge in the latter two charts. The top 100 websites are simply more predisposed to strongly oppose the trends created by other entries, and this is most likely a result of their more advanced and optimized construction.

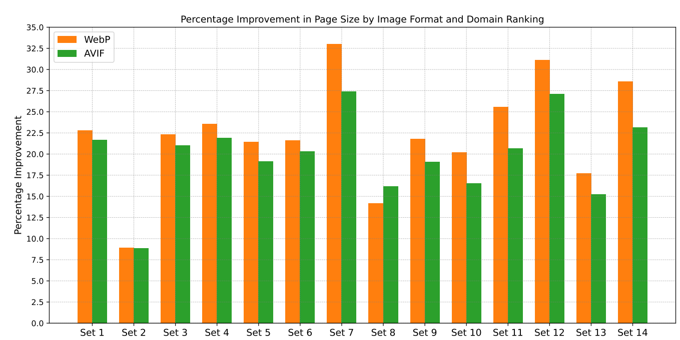

Figure 9: Percentage Change in Page Size, by popularity

4.3 Image-Level Analysis

With the Income and Representative Set analyses out of the way, we can now look at our large image cache to extract some insights uncoupled from the URLs. We split our set of images into deciles by image size, and make 3 charts:

- Average Percentage gains offered by WebP/AVIF vs JPG/PNG both (Figure 12)

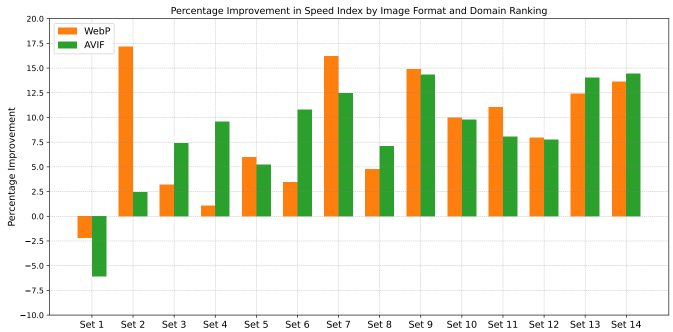

Figure 10: Percentage Change in Speed Index, by popularity

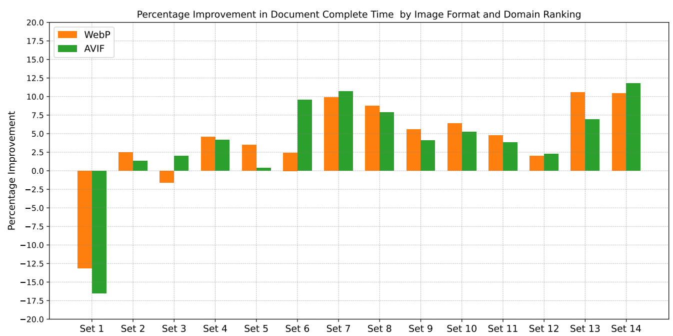

Figure 11: Percentage Change in Document Complete Time, by popularity

- Average Percentage gains offered by WebP/AVIF vs JPG only (Figure 13)
- Average Percentage gains offered by WebP/AVIF vs PNG only (Figure 14)

The insights here are quite interesting, and are summed up as follows:

- AVIF reductions are highly correlated to image size, while WebP is somewhat correlated and more reliably creates greater reductions for a wide range of image sizes.
- AVIF is better at reducing the size of JPGs than it is for PNGs, and WebP is better at both.
- Somewhere in between the 2nd and 3rd decile of PNG image sizes, AVIF reductions run into an inflection point. PNG images under 1.6 KB (which is the upper limit of the second decile) tend to gain size if converted to AVIF, and this effect increases in the first decile. WebP performs as expected and delivers at least a 20% reduction in image size regardless of image size, even in the first decile.

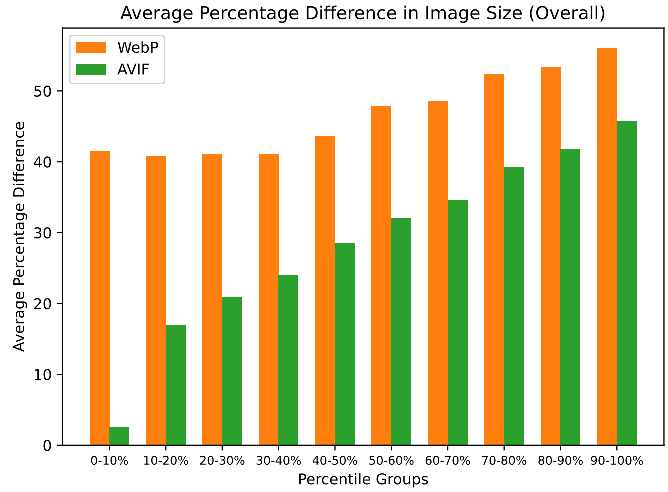

Figure 12: Percentage Reduction in Image Size, WebP/AVIF vs Legacy Formats overall, by deciles

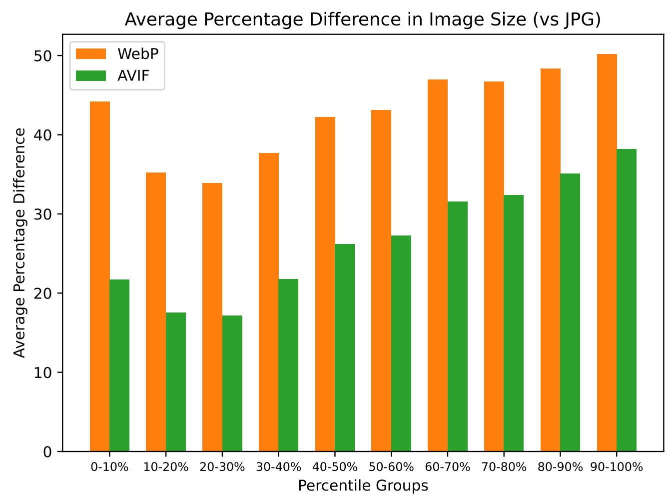

Figure 13: Percentage Reduction in Image Size, WebP/AVIF vs JPG, by deciles

5 CONCLUSION

In this analysis, we delve into the macro effects of employing WebP and AVIF formats instead of the conventional PNG/JPG formats. We focus on various key metrics including page size, page load time, First Contentful Paint (FCP), Speed Index, and Document Complete Time.

Our results indicate that the use of WebP and AVIF leads to considerable reductions in page size for both high and low-income groups. While the mean values exhibit a decrease of about 25% for high-income countries and 35% for low-income countries, the median values do not show a similar substantial improvement. This points to the fact that larger

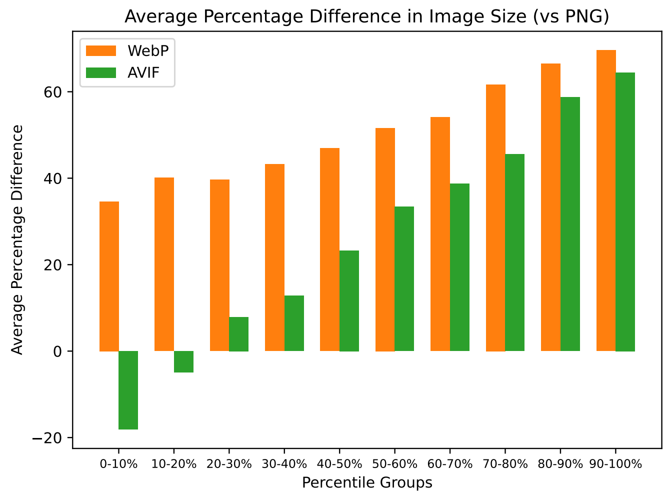

Figure 14: Percentage Reduction in Image Size, WebP/AVIF vs PNG by deciles

pages are more likely to benefit from the switch to WebP and AVIF in terms of image size. Furthermore, our analysis of the Speed Index metric reveals a promising picture. The application of WebP and AVIF significantly influences a web page's Speed Index, with more than 80% of the webpages experiencing a reduction in Speed Index for both formats across all income groups. When it comes to the First Contentful Paint (FCP) metric, our findings are equally encouraging. Both high and low-income groups show improvements, with the low-income group reaping slightly better results.

However, the picture is not as rosy when we examine the Page Load Time and Document Complete Time metrics. Despite seeing improvements, these are somewhat subdued for larger pages across income groups. Still, over 80% of web pages experienced a net reduction in page load time.

In the Representative Set Analysis, a trend emerged where the top 100 ranked websites generally perform worse when switched to WebP or AVIF. This could be attributed to their high levels of prior optimization. Our Image-Level Analysis provided some interesting insights as well, and by splitting our set of images into deciles by image size we found that the AVIF reductions are highly correlated to image size, whereas WebP reductions are generally consistent across the image sizes and larger than AVIF reductions.

Overall, our analysis presents a compelling case for the adoption of WebP and AVIF formats. While they may not be suitable for all web pages, they can significantly enhance web performance, particularly for larger pages and those in low-income countries. Switching to these formats represents a low-effort optimization that can make web pages from Low Income countries significantly more agile and competitive, and will have a great effect on bandwidth conservation and network de-congestion along with an enhanced user experience.

6 AREAS OF FURTHER RESEARCH

While our current study provides substantial insights, there are additional areas of research that can be explored to more conclusively establish the performance of these next-generation image formats. A category-wise analysis, for instance, could offer a deeper understanding of the impact of WebP and AVIF on different types of content, such as news versus sports. Additionally, a comparative study on the processing power and time required to decode WebP, AVIF, and JPG/PNG images could provide a more comprehensive view of the computational resources needed for these formats, especially in the context of mobile devices where power efficiency is key. Another valuable avenue for future research could involve repeating our analysis on a high-end mobile device, as this would shed light on how these image formats perform in a resource-abundant environment, and if the gains on offer are better. Finally, considering the varying network conditions and types, conducting the analysis under different network scenarios could offer a more realistic view of the practical implications of these formats.

Each of these further research directions could contribute to a more holistic and definitive understanding of the performance of WebP and AVIF.

REFERENCES

[1] [n.d.]. Google Next Billion users - research. https://nextbillionusers.google/research/
[2] 2021. Mobile data costs fall but as demand for internet services surges, progress remains too slow. https://webfoundation.org/2021/03/mobile-data-costs-fall-but-as-demand-for-internet-services-surges-progress-remains-too-slow/
[3] 2022. Page weight. https://httparchive.org/reports/page-weight
[4] Nabajeet Barman and Maria G. Martini. 2020. An Evaluation of the Next-Generation Image Coding Standard AVIF. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–4. https://doi.org/10.1109/QoMEX48832.2020.9123131
[5] Netflix Technology Blog. 2020. AVIF for next-generation image coding. https://netflixtechblog.com/avif-for-next-generation-image-coding-b1d75675fe4
[6] Giaime Ginesu, Maurizio Pintus, and Daniele D. Giusto. 2012. Objective assessment of the WebP image coding algorithm. *Signal Processing: Image Communication* 27, 8 (2012), 867–874. https://doi.org/10.1016/j.image.2012.01.011 Special issue on: pervasive mobilemultimedia.
[7] Zhanjun Si and Ke Shen. 2016. Research on the WebP Image Format. In *Advanced Graphic Communications, Packaging Technology and Materials*, Yun Ouyang, Min Xu, Li Yang, and Yujie Ouyang (Eds.). Springer Singapore, Singapore, 271–277.