Real Time Arbitrary Video Style Transfer

Outline

1. Problem Overview
2. Proposed Model Architecture
3. Learning Process
4. Experiments
5. Progress
6. Challenges
7. Next Steps
8. References

1.

Problem Overview

Video Style Transfer

Video Style Transfer

Transferring a *Style* extracted from an image into a whole video sequence

Each frame is processed and a new frame is synthesized with the content of the frame and the style of the image

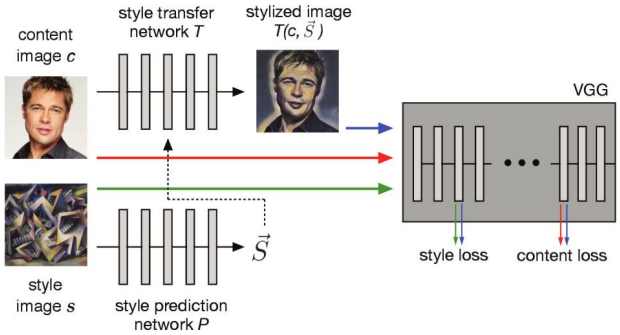
2.

Proposed Model Architecture

Arbitrary Style Transfer

The architecture comprises of two sub networks

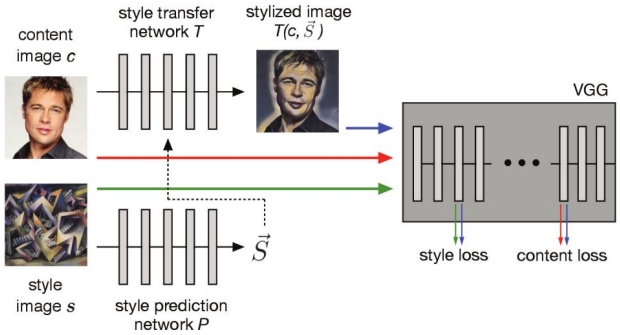
- The Style Prediction Network
- The Image Transformation Network



Arbitrary Style Transfer

The style image is fed to the network *P* to produce the vector *S*

The vector *S* is fed to the Network *T* along with the content image to produce the stylized image



The Transform Network

It’s essentially an auto-encoder with a normalization function on the output

3 Conv layers, 5 Residual Blocks (Encoder)

2 Upsampling, 1 Conv layer (Decoder)

The Transform Network

The output is normalized

$$\tilde{z} = \gamma_s \left(\frac{z - \mu}{\sigma} \right) + \beta_s$$

Where *Gamma* and *Beta* are the vector *S* produced by *P*

The Transform Network

The intuition here is that the styles share common features and images can be mapped from a style to another using linear transformation on the feature map.

The Style Network

To produce the vector *S*, the network is fed the style image and produces the vector.

This takes into account the shared features and textures between different styles.

3.

Learning Process

Image Arbitrary Style Transfer

Using 80,000 painting to train the style/transform network

Comparable results on unobserved styles

Video Arbitrary Style Transfer

To train the previous network for video style transfer,

It must preserve temporal consistency between frames

Two frame learning synergic

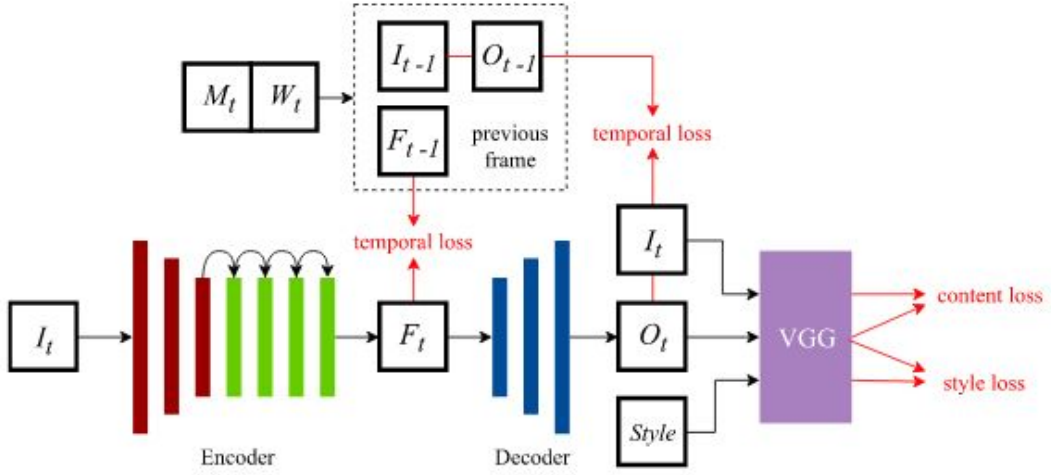
1. Given a dataset of videos, group each two consecutive frames together.
2. Pick a pair at random and perform forward pass in the network for each frame.
3. Compute the losses using

$$\mathcal{L}(t-1,t) = \sum_{i \in \{t-1,t\}} (\alpha \mathcal{L}_{content}(i) + \beta \mathcal{L}_{style}(i) + \gamma \mathcal{L}_{tv}(i)) + \lambda_f \mathcal{L}_{temp,f}(t-1,t) + \lambda_o \mathcal{L}_{temp,o}(t-1,t)$$

1. Update network parameters

Two frame learning synergic

Uses multiple level temporal losses at the output and at the input



Losses

Content Loss

L2 distance between the feature maps of VGG of input and output

Style Loss

L2 distance between the gram matrices of input and output

Temporal Loss (feature maps)

Masked L2 distance between 2nd feature map and the warp of 1st using ground truth optical flow

Temporal Loss (output maps)

Same as feature maps, but on output layer and distance in illuminance channel

Network initialization

Network is initialized with the image arbitrary style transfer.

The network has already optimized content and style loss, good starting point.

4.

Experiments

Two approaches

To solve the real-time arbitrary video style transfer, two approaches are proposed

1. Frame by Frame inference
2. Enforcing Temporal Consistency on Image arbitrary style transfer

Frame by Frame inference

We conjectured that since the network perform styling using forward pass using high-level features

Objects in consecutive frames will have same representation in the output as opposed to methods used by Ruder[1]

1. A similar concept was discussed by Huang et al [4]

Enforcing Temporal Consistency on Image arbitrary style transfer

Using knowledge transfer, the image arbitrary style network is trained to minimize a new objective function initialized to published weights

The training process is adopted from that of the ReCoNet in [5]

Enforcing Temporal Consistency on Image arbitrary style transfer

The network starts off while it’s already minimized the style and content loss

It then tries to minimize temporal loss while keeping content and style loss minimal

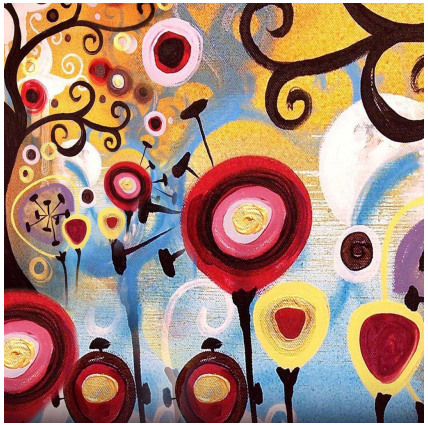
The style embedding network is frozen while the transformer model is trained

Enforcing Temporal Consistency on Image arbitrary style transfer

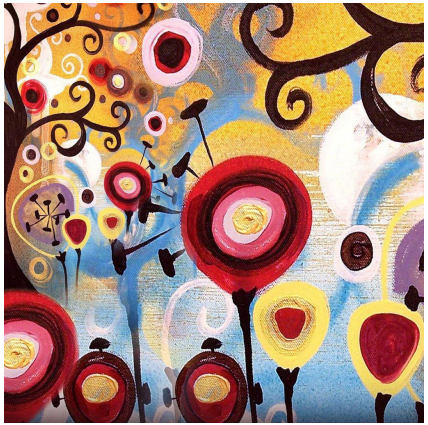
Care must be taken with training examples to train variety of painting styles as was recommended by Ghiasi in [6]

5.

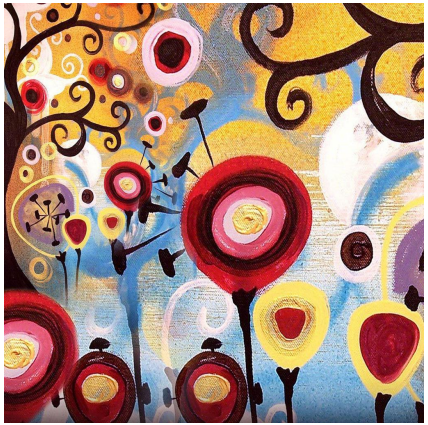
Progress



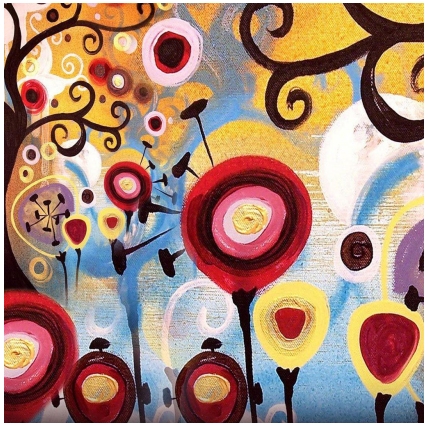
Bandage 2 + Candy



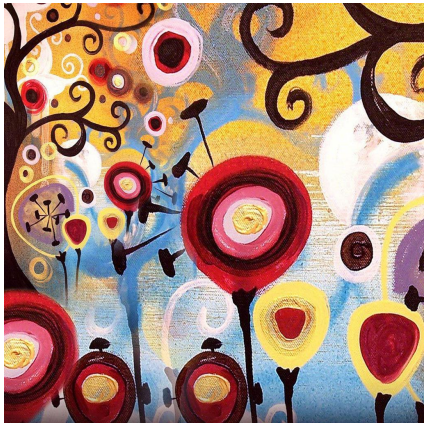
Alley 2 + Candy



Ambush 5 + Candy



Temple 2 + Candy



Market 6 + Candy

Approach #1

- Quantitative analysis

  Stability error is the square root of output-level temporal error over one whole scene

$$e_{stab} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\frac{1}{D}M_t\|O_t - W_t(O_{t-1})\|^2}$$

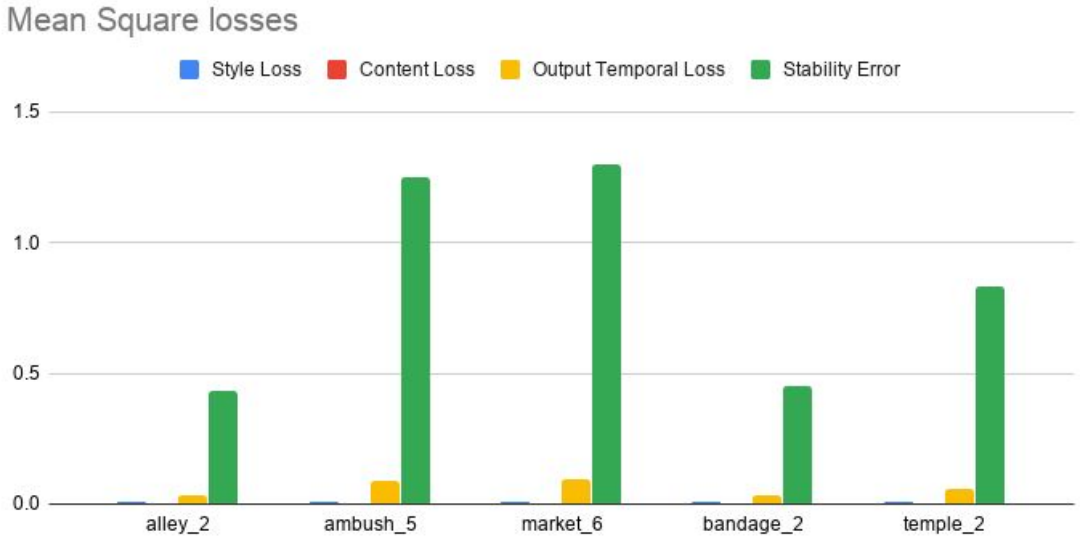
Approach #1

	Style Loss	Content Loss	Output Temporal Loss	Variation Regulizer	Stability Error
alley_2	0.0116	0.0001	0.0312	43.4142	0.4324
ambush_5	0.0117	0.0001	0.0900	42.9696	1.2489
market_6	0.0115	0.0001	0.0935	43.9220	1.2970
bandage_2	0.0116	0.0001	0.0327	43.7613	0.4534
temple_2	0.0115	0.0001	0.0597	43.4582	0.8296

Model	Alley-2	Ambush-5	Bandage-2	Market-6	Temple-2
Chen *et al* [4]	0.0934	0.1352	0.0715	0.1030	0.1094
ReCoNet	0.0846	0.0819	0.0662	0.0862	0.0831
Huang *et al* [17]	0.0439	0.0675	0.0304	0.0553	0.0513
Ruder *et al* [27]	0.0252	0.0512	0.0195	0.0407	0.0361

Comparison Between State-of-the-art models in Video Transfer and our current model

Approach #1



Total Loss factors

$$e_{stab} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\frac{1}{D}M_t\|O_t - W_t(O_{t-1})\|^2}$$

Conclusion

- The model has low style & content loss - as expected
- Although the temporal loss is low, it’s still one order of magnitude higher.
  - The model will benefit from training using the approach discussed

Approach #2

- This approach continue the training on the pre-trained model to minimize the temporal loss as discussed
- The migration from Pytorch to TF and downgrade is in progress.
- There have been compatibility issues as discussed in the next sections, which are almost solved.

Approach #2

- The training is expected to take long time
  - 2-frame synergic means batch size of 2!
  - Large number of frames to train on.
  - For all frames, we must train on multiple styles
  - And finally, repeat for a suitable number of epochs!
- Colab code here

*However it's expected to take small number of epochs, since 2 out of the 4 losses are already optimal, and 1 close to optimal

6.

Challenges

Code Incompatibilities

- The arbitrary style transfer code was implemented as part of Magenta project
- The code is implemented using TF v1, using many deprecated calls and even deleted libraries from TF v2
- The ReCoNet is implemented in Torch

Code Incompatibilities

- The integration of both caused a lot of issues
  - Different dimensions
  - Different API
  - Deprecated calls
- Some Functions didn’t even exist in TF v1 as the warp function necessary for computing temporal losses

7.

Next Steps

Approach #2

- Continue the training loop and the training.
- Convert the model to TFLite format to produce real-time inference.
- Compare the results against Approach #1 and state-of-the-art-models again.

8.

References

References

- [1] L. A. Gatys, et al., “A neural algorithm of artistic style,” 2015.
- [2] J. Johnson, A. Alahi, et al., “Perceptual losses for real-time style transfer and super-resolution,” 2016.
- [3] M. Ruder, et al., “Artistic style transfer for videos,”
- [4] H. Huang, et al., “Real-time neural style transfer for videos,”, July 2017.
- [5] C. Gao, et al., “Reconet: Real-time coherent video style transfer network,” 2018.
- [6] G. Ghiasi, et al., “Exploring the structure of a real-time, arbitrary neural artistic stylization network,” 2017

Thanks!

Any questions?

Quotations are commonly printed as a means of inspiration and to invoke philosophical thoughts from the reader.

This is a slide title

- Here you have a list of items
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content, but won’t do both.

Big concept

Bring the attention of your audience over a key concept using icons or illustrations

You can also split your content

White

Is the color of milk and fresh snow, the color produced by the combination of all the colors of the visible spectrum.

Black

Is the color of coal, ebony, and of outer space. It is the darkest color, the result of the absence of or complete absorption of light.

In two or three columns

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

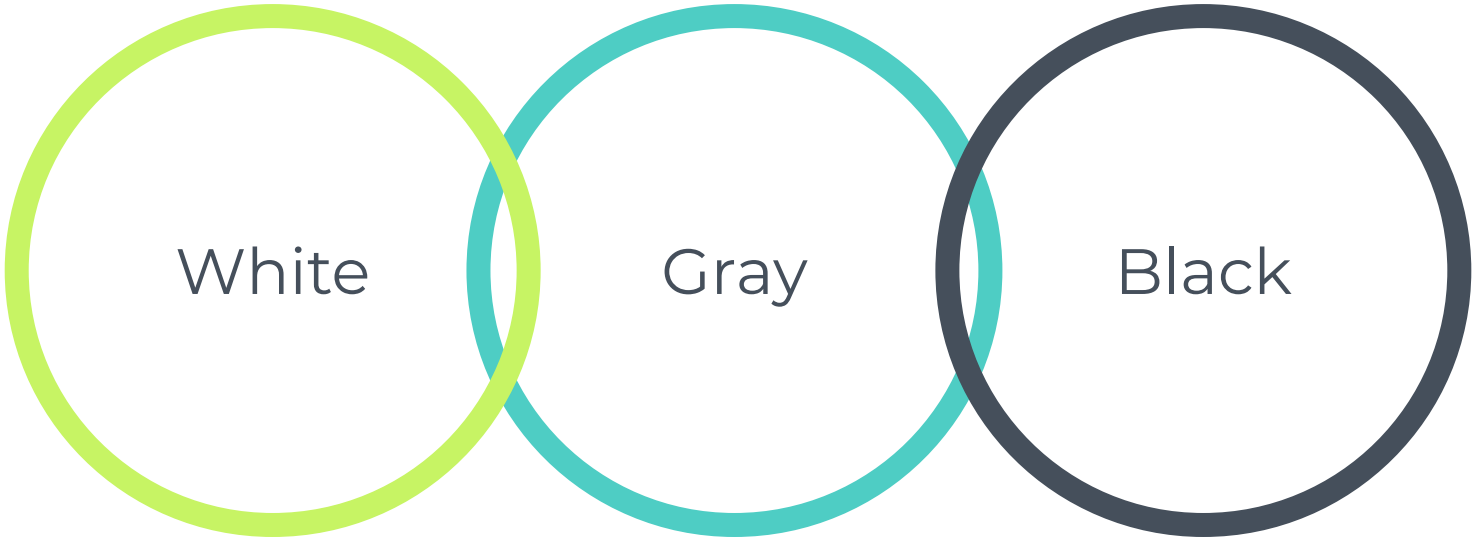
A picture is worth a thousand words

A complex idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.

Want big impact?

Use big image.

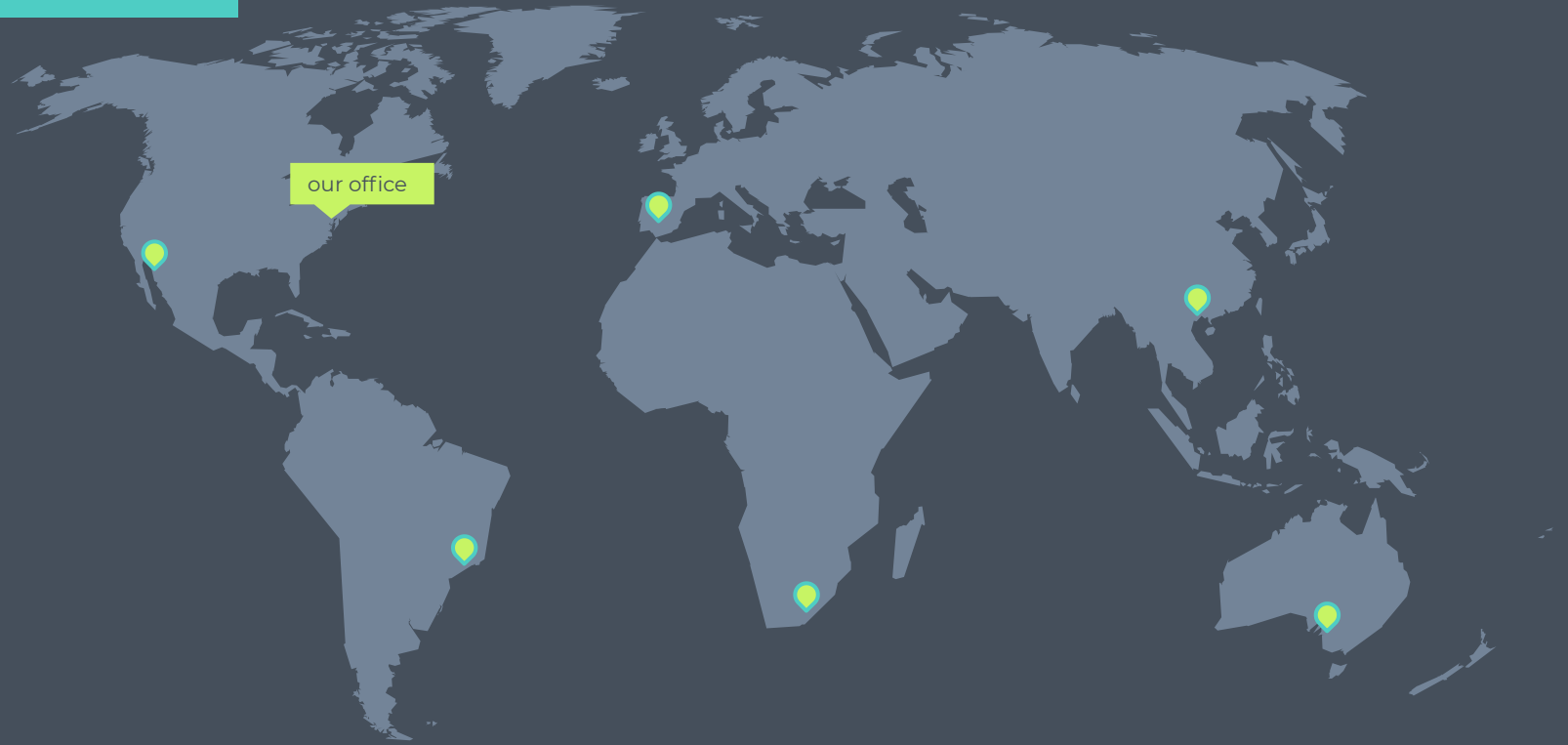
Use charts to explain your ideas



And tables to compare data

	A	B	C
Yellow	**10**	**20**	**7**
Blue	**30**	**15**	**10**
Orange	**5**	**24**	**16**

Maps



89,526,124

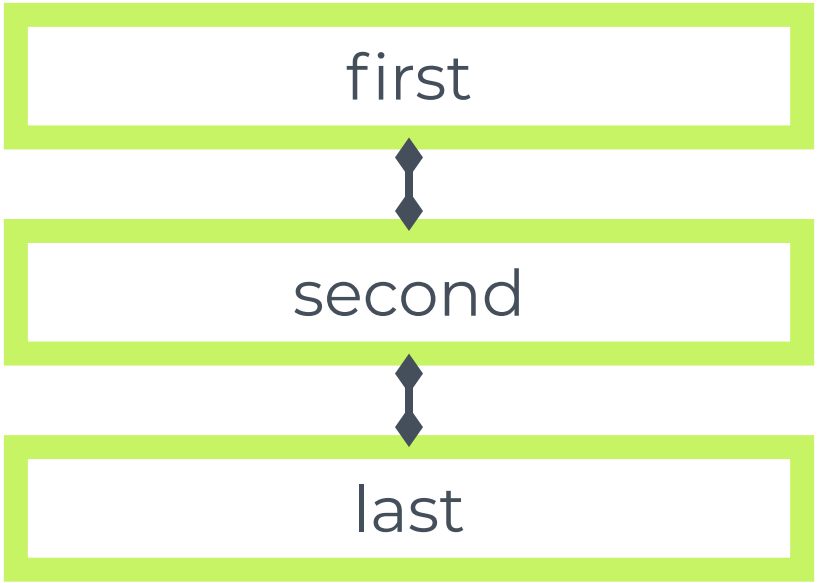
Whoa! That’s a big number, aren’t you proud?

89,526,124$
That’s a lot of money

185,244 users
And a lot of users

100%
Total success!

Our process is easy



Let’s review some concepts

Yellow
Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

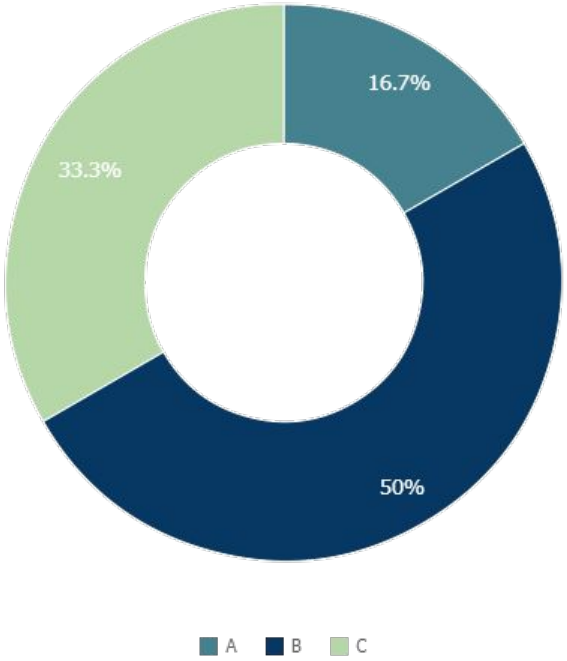
Blue
Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

Yellow
Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue
Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

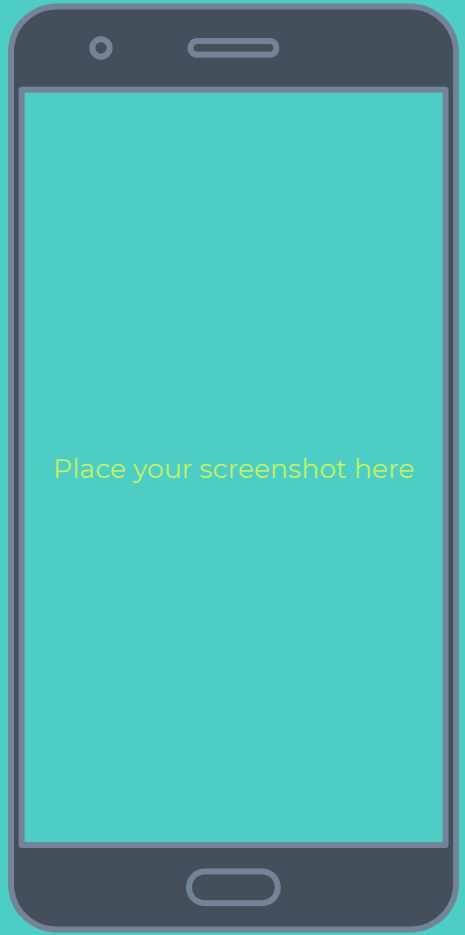
Red
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



You can copy&paste graphs from Google Sheets

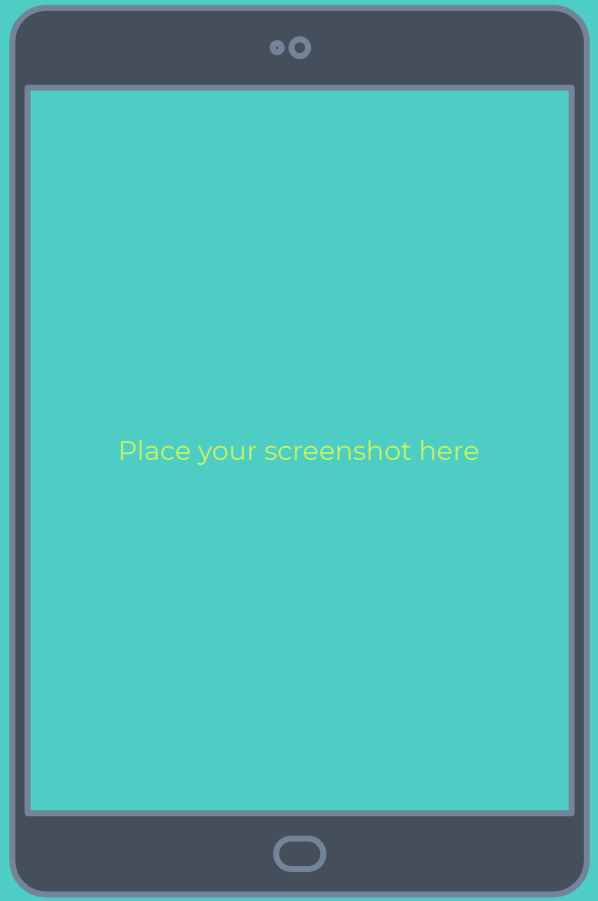
Mobile project

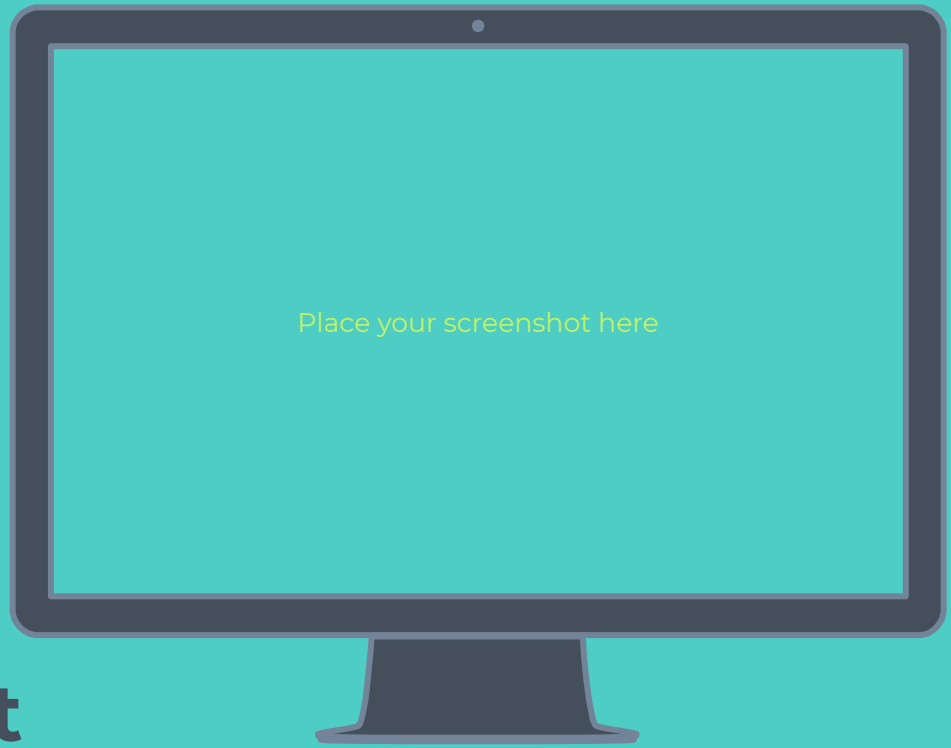
Show and explain your web, app or software projects using these gadget templates.



Tablet project

Show and explain your web, app or software projects using these gadget templates.





Desktop project

Show and explain your web, app or software projects using these gadget templates.

Presentation design

This presentations uses the following typographies and colors:

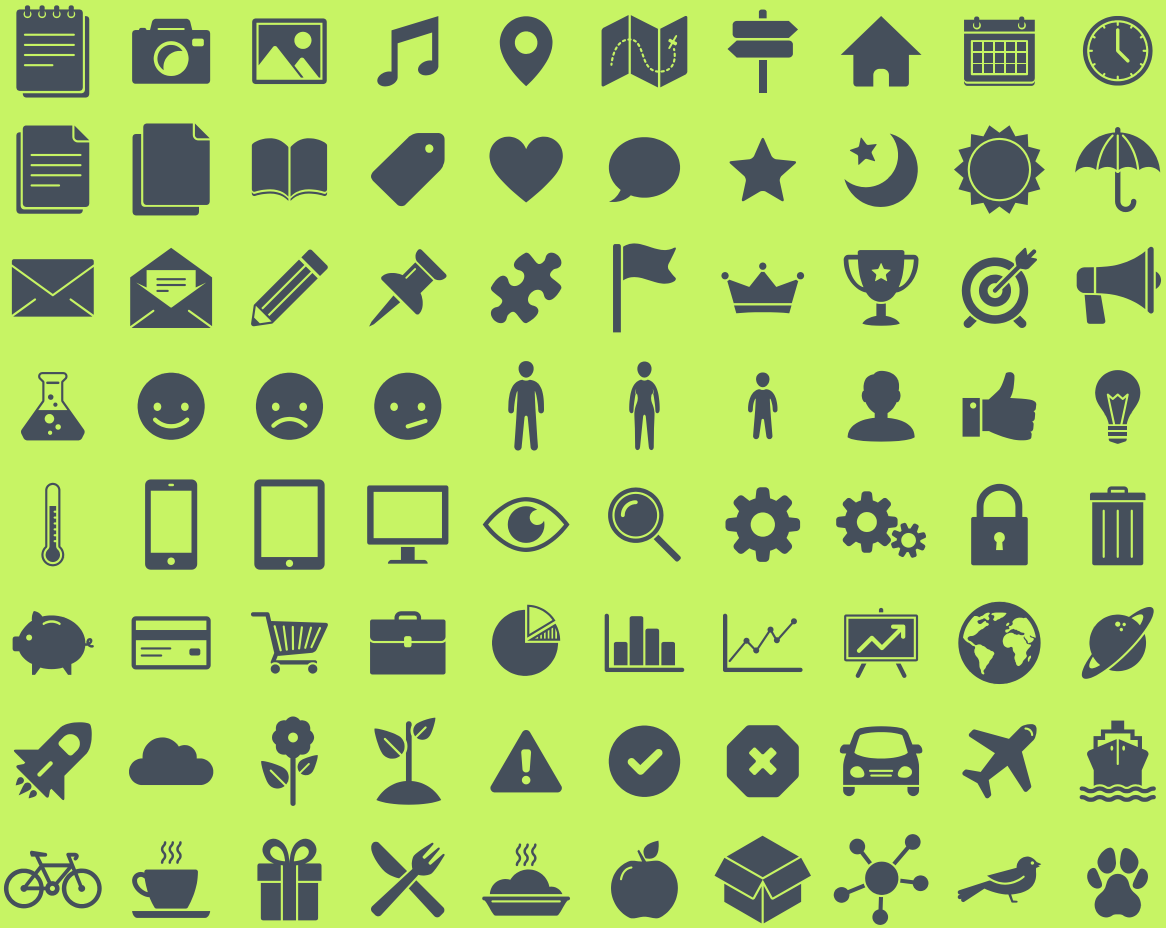
- Titles & Body copy: **Montserrat**

You can download the fonts on this page:

https://www.fontsquirrel.com/fonts/montserrat

- Grey **#454f5b**
- Light grey **#738498**
- Neon green **#c7f464**
- Aqua **#4ecdc4**

You don’t need to keep this slide in your presentation. It’s only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®



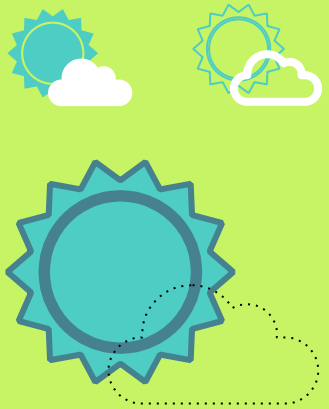
SlidesCarnival icons are editable shapes.

This means that you can:

- Resize them without losing quality.
- Change fill color and opacity.
- Change line color, width and style.

Isn't that nice? :)

Examples:



Now you can use any emoji as an icon!
And of course it resizes without losing quality and you can change the color.

How? Follow Google instructions
https://twitter.com/googledocs/status/730087240156643328

and many more...

Slides Carnival

Free templates for all your presentation needs

For PowerPoint and Google Slides

100% free for personal or commercial use

Ready to use, professional and customizable

Blow your audience away with attractive visuals

Instructions for use

EDIT IN GOOGLE SLIDES

Click on the button under the presentation preview that says "Use as Google Slides Theme".

You will get a copy of this document on your Google Drive and will be able to edit, add or delete slides.

You have to be signed in to your Google account.

EDIT IN POWERPOINT®

Click on the button under the presentation preview that says "Download as PowerPoint template". You will get a .pptx file that you can edit in PowerPoint.

Remember to download and install the fonts used in this presentation (you'll find the links to the font files needed in the Presentation design slide)

More info on how to use this template at www.slidescarnival.com/help-use-presentation-template

Hello!

I am Jayden Smith

I am here because I love to give presentations.
You can find me at @username