

開源軟體介紹

- 以OPENCV為例

葉信和 / Hsin-Ho Yeh

Founder / CEO / Software Engineer @ 信誠金融科技

hsinho.yeh@footprint-ai.com



Tintin

Machine Learning Platform for Everyone



信誠金融科技股份有限公司
XINCHEN FINTECH CO., LTD.

2023/10



Download Slides

<http://bit.ly/3QmqsTw>

<https://github.com/FootprintAI/talks/tree/main/slides>



Intro & timeline



- footprint-ai.com (信誠金融科技) is committed to providing Software-as-a-Service for serving AI/ML applications, specializing in:
 - Cloud-native architecture, MLOps, Green Tech, Internet-scale Data Analysis
- Practical experience in taking an idea (zero) to many products (one)
 - Singapore-based insurtech startup iglooinsure (18M USD/2020 Series A, 46M USD/2023 Series B).
- Business partner with IBM Taiwan (IBM.tw)
- Joined Nvidia Inception program in 2021.
- Moved into FinTechSpaces in 2023.

About me

- 2020 - Present at 信誠金融科技
 - Tintin: a machine learning platform for everyone
 - <https://get-tintin.footprint-ai.com>
- 2016 - 2020 at IglooInsure (16M+ in series A+ 2020)
 - Provide digital insurance for e-economic world
 - Funded in KUL, Headquartered in Singapore
 - First employee/ Engineering Lead / Regional Head/ Chief Engineer
- 2013 - 2016 at Studio Engineering @ hTC
 - Principal Engineer on Cloud Infrastructure Team
- 2009 - 2012 at IIS @ Academia Sinica
 - Computer vision, pattern recognition, and data mining
- CS@CCU, CS@NCKU alumni





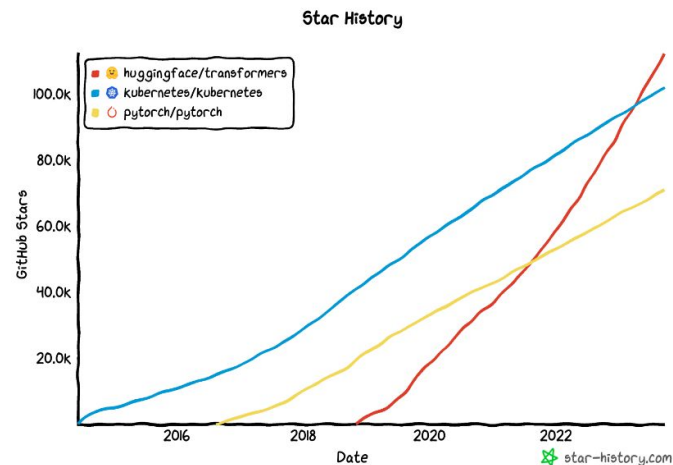
Agenda

- What is Open Source?
- OpenCV Introduction
- How to contribute to Open Source?
- OpenCV Case Study
- FAQ

Huggingface

- An open source data science and machine learning platform, aka github for AI.
- Notable with its open source transformer library, which simplifies the process of working with pre-trained models.
- Huggingface Hub [1]
 - A central hub for community to share and explore machine learning.

year	2021		2023
Fundraising status	40M (B)		235M(D)



https://en.wikipedia.org/wiki/Hugging_Face
<https://github.com/huggingface/transformers>
<https://collabnix.com/what-is-hugging-face-and-why-it-is-damn-popular>



What is open source competitive advantage?

- Open Source (OS) company has gone to public: Hashicorp, JFrog, Elastic, MongoDB and Gitlab
- Key advantages from open source
 - Crowdsourcing product development
 - anyone on the internet can contribute.
 - low-cost sales channel: anyone can download / use the software for free.
 - Trust to the provider is increasing during daily based usage.
 - Acquire a base of free users and offer them with upselling.



What is Open Source Software?

“Open-source software is a type of computer software ... the copyright holder grants users the rights to **use**, **study**, **change**, and **distribute** the software to anyone and for any purpose.” [1]

[1] https://en.wikipedia.org/wiki/Open-source_software

What is Open Source Software? (1/4)



	Open Source Software	Proprietary Software
Ownership	Everyone	A group of limited people
Usability	Difficult use for unskilled users Ex: Desktop Linux	Feature neat to end-user Ex: Microsoft Windows
Security	Better (? Ex: <u>PHP source code is injected backdoors by hackers</u>	Less
Stability	Less	Better
Opacity	Transparent	Closed

Ref: <http://www.optimusinfo.com/downloads/white-paper/open-source-vs-proprietary-software-pros-and-cons.pdf>

What is Open Source Software? (2/4)



Misunderstanding on open source software

- Open source software is **Free** in terms of Freedom, not Monetary.
- Open source developer is **NOT** a free labor [1]
 - Sponsorship by companies
 - Add values(a.k.a layers) on the existing project
 - Professional consult
- Open source means **Pool** quality
 - Good quality when the project gained enough eyeballs
 - On average the density of code defect is smaller than proprietary software [2]

[1] <https://softwareengineering.stackexchange.com/questions/100685/making-money-with-open-source-as-a-developer>

[2] <https://blogs.worldbank.org/opendata/quality-open-source-software-how-many-eyes-are-enough>

What is Open Source Software? (3/4)

Where open source communities lives?

- Github
 - Free for any public repositories
 - Github Sponsor

Become a sponsor to **OpenCV**



OpenCV is the largest Computer Vision library in the world. It's used worldwide: in the world's biggest companies, underwater, and in space.

OpenCV was started in 1998 and played the key role in the development of Computer Vision and Artificial Intelligence. It is released under Apache 2.0 license and is open-source and free for everyone.

In the sponsorship program, our main focus is direct support of OpenCV core developers. This means more new functionality, better support, and faster code review for the pull requests from our huge community.

Support OpenCV today to help us develop Computer Vision industry worldwide.

Current sponsors 3



Past sponsors 5



What is Open Source Software? (4/4)



Key files to understand a github project

- README.md
 - How to build and use the project
- LICENSE
 - Terms and Conditions
- CONTRIBUTING.md
 - Rules on contributing

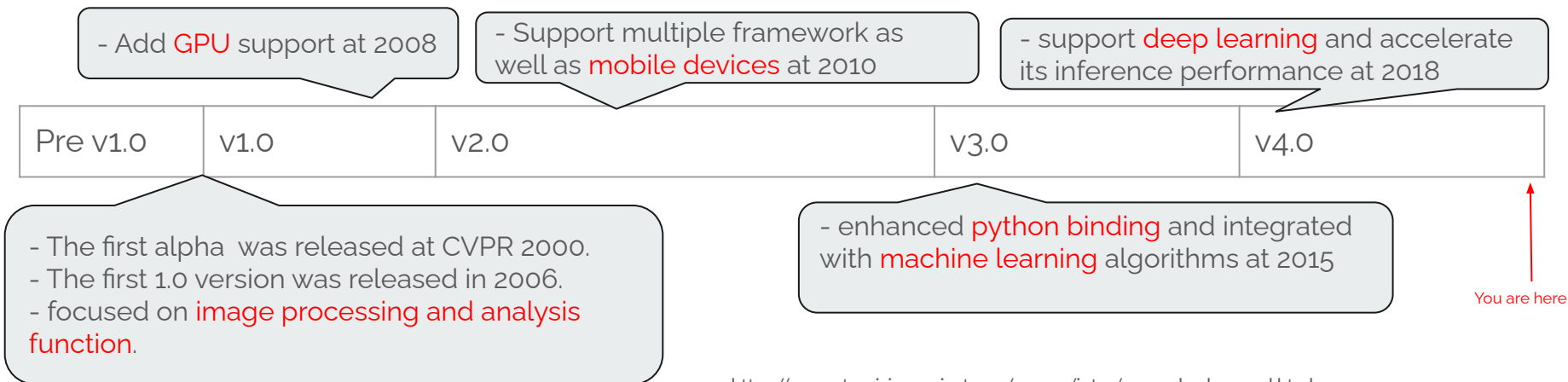


OpenCV

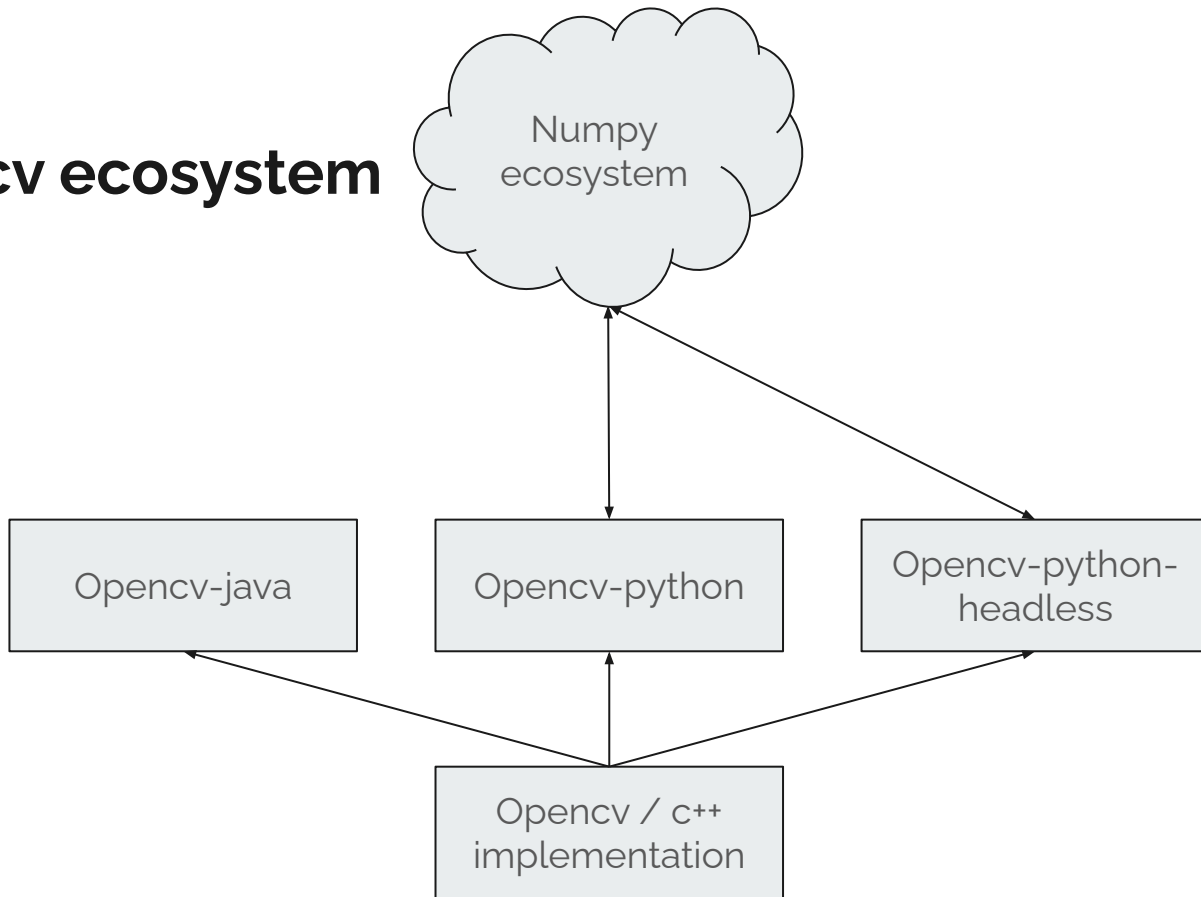
introduction

Opencv history

OpenCV (Open Source Computer Vision Library) is an open-source computer vision and machine learning software library.



Opencv ecosystem





Opencv installation

```
// pre-build binaries  
https://github.com/opencv/opencv/releases/tag/4.8.1
```

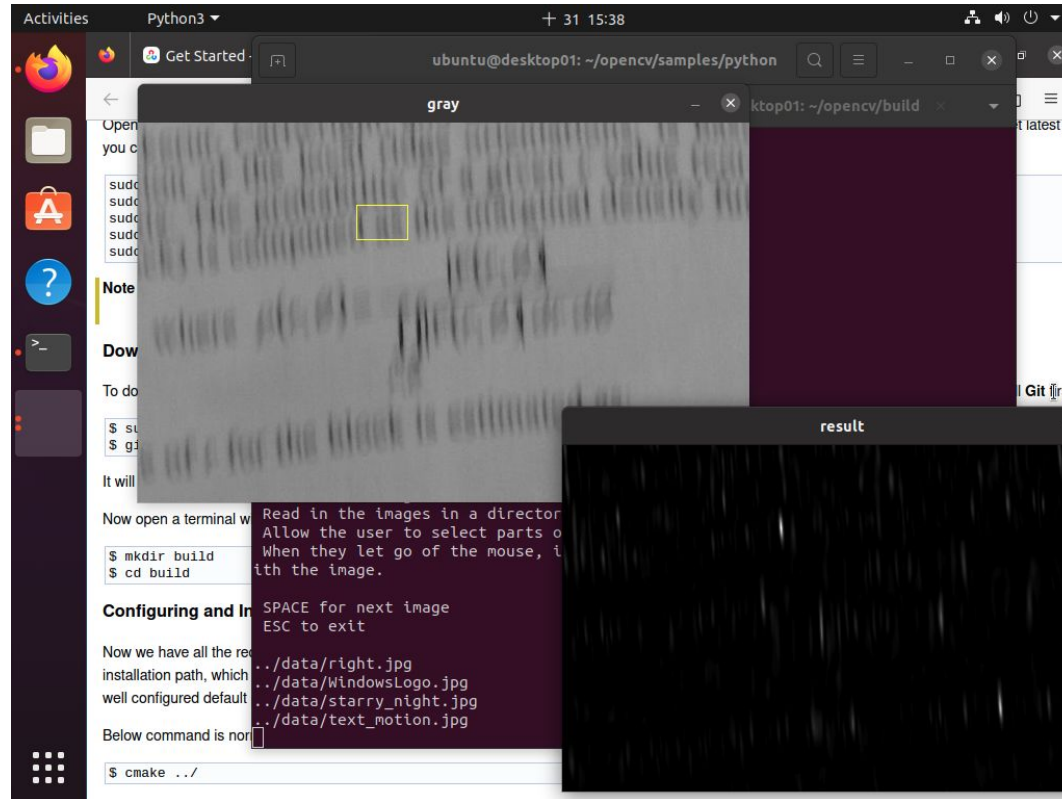
```
// python pypi (python package index)  
pip3 install opencv-python
```



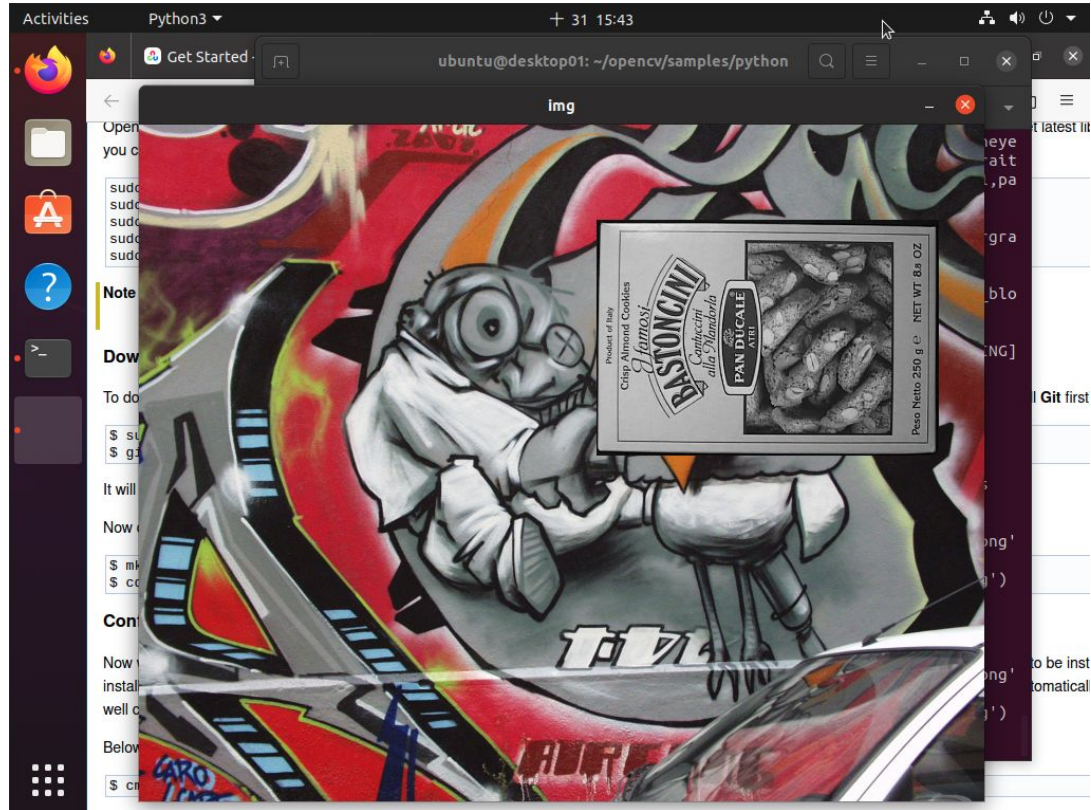

Opencv: installation from source code (advanced)

- Build with cpp
 - Linux: https://docs.opencv.org/master/d7/d9f/tutorial_linux_install.html
 - Windows:
https://docs.opencv.org/master/d3/d52/tutorial_windows_install.html
- Build with python
 - Linux:
https://docs.opencv.org/4.x/d2/de6/tutorial_py_setup_in_ubuntu.html
 - Windows:
https://docs.opencv.org/4.x/d5/de5/tutorial_py_setup_in_windows.html
- Took ~~several minutes~~ half hours to build code

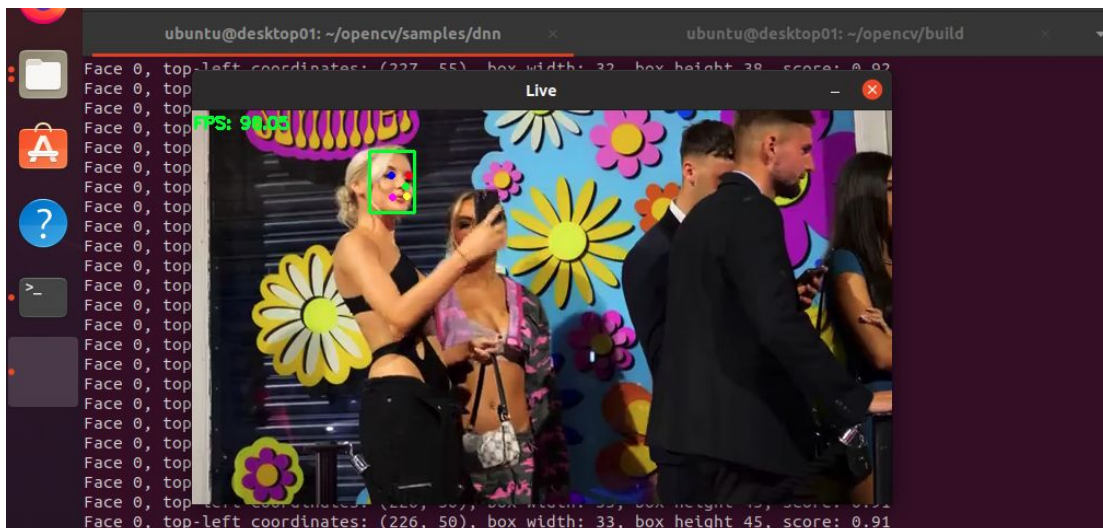
Example I mouse_and_match



Example II scene_render

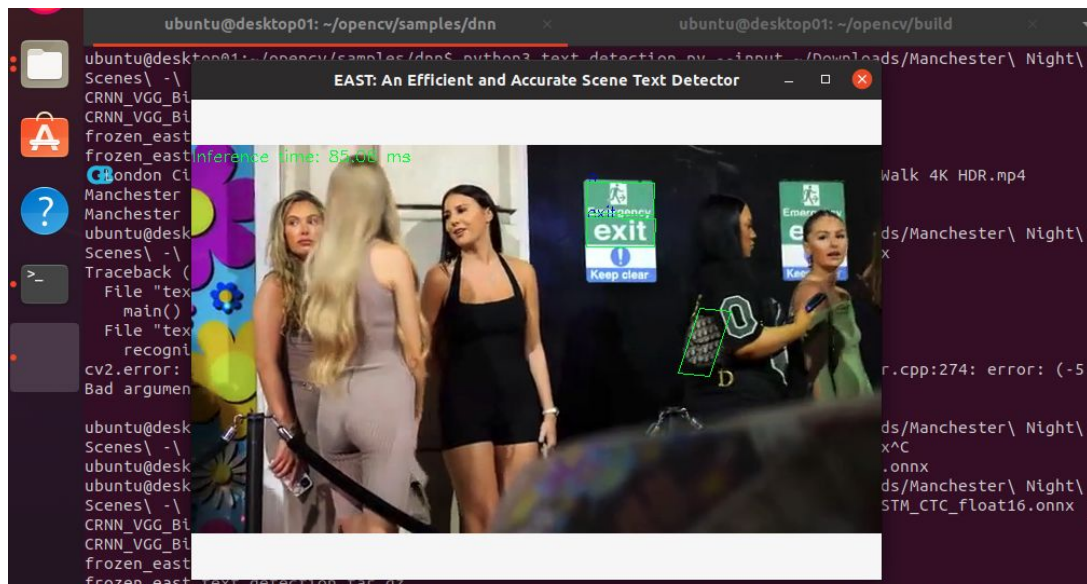


Example III face detector



Video source: <https://www.youtube.com/watch?v=ufPn6RaDxbg>

Example III text detector



Video source: <https://www.youtube.com/watch?v=ufPn6RaDxbg>



How to contribute Open Source?

How to contribute to open source?



The two most popular way to contribute

1. Raising an issue where you think the project could be improved
 - a. Report a problem/bug (kind/bug)
 - b. Suggest new features (kind/feature)
 - c. Gap in documentation (doc)
2. Contribute to code
 - a. Pull-Request in Github

🔔

415 Open

✓ 1,720 Closed

Author ▾

Label ▾

Projects ▾

Milestones ▾

Assignee ▾

Sort ▾

🔔

[frontend] The sort, filter of artifact does not work

area/frontend

kind/bug

#5565 opened 20 hours ago by capri-xiyue

1

🔔

[Doc] Update "Visualize Results in the Pipelines UI" page

#5564 opened yesterday by Bobqv

2

🔔

[backend] OutputArtifact path is None

area/backend

kind/bug

#5560 opened 2 days ago by wilbry

🔔

[feature] Configurable KFP Artifact types while storing to Minio

area/sdk

kind/feature

#5558 opened 2 days ago by rado-narrak

2

🔔

[Multi User] Kubeflow Pipelines Missing Headers

area/backend

kind/bug

#5556 opened 3 days ago by DeyverRamirez

🔔

[feature] Allow jobs to be scheduled on AWS Fargate

area/backend

kind/feature

#5555 opened 3 days ago by yuhuishi-convect

2

🔔

[sdk] mlpipelines-metrics path is wrong when using kfp.components.load_component_from_text

area/sdk

kind/bug

#5548 opened 4 days ago by santiagoolivar2017

🔔

pipeline_conf.set_ttl_seconds_after_finished(seconds=10) NOT WORKING

area/backend

kind/bug

#5545 opened 4 days ago by prasadkyp7

🔔

[sdk] input parameter created without default value

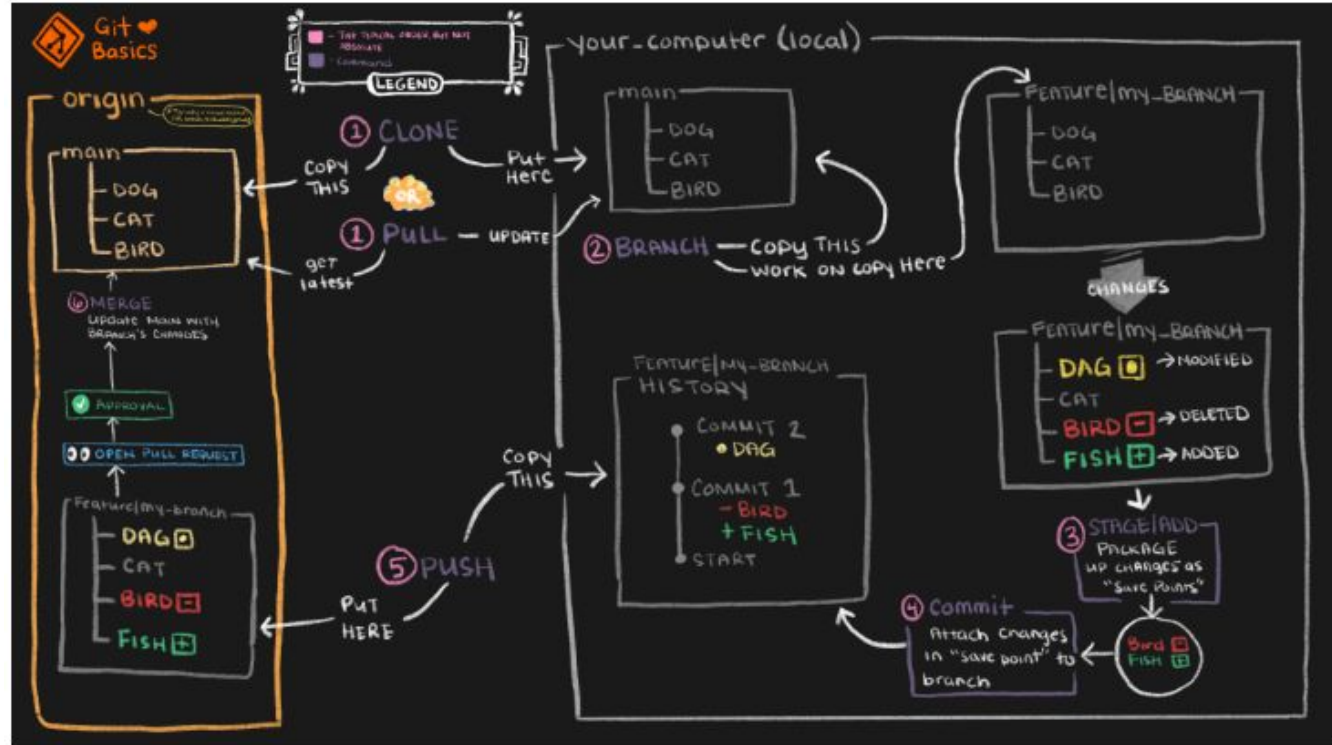
area/sdk

kind/bug

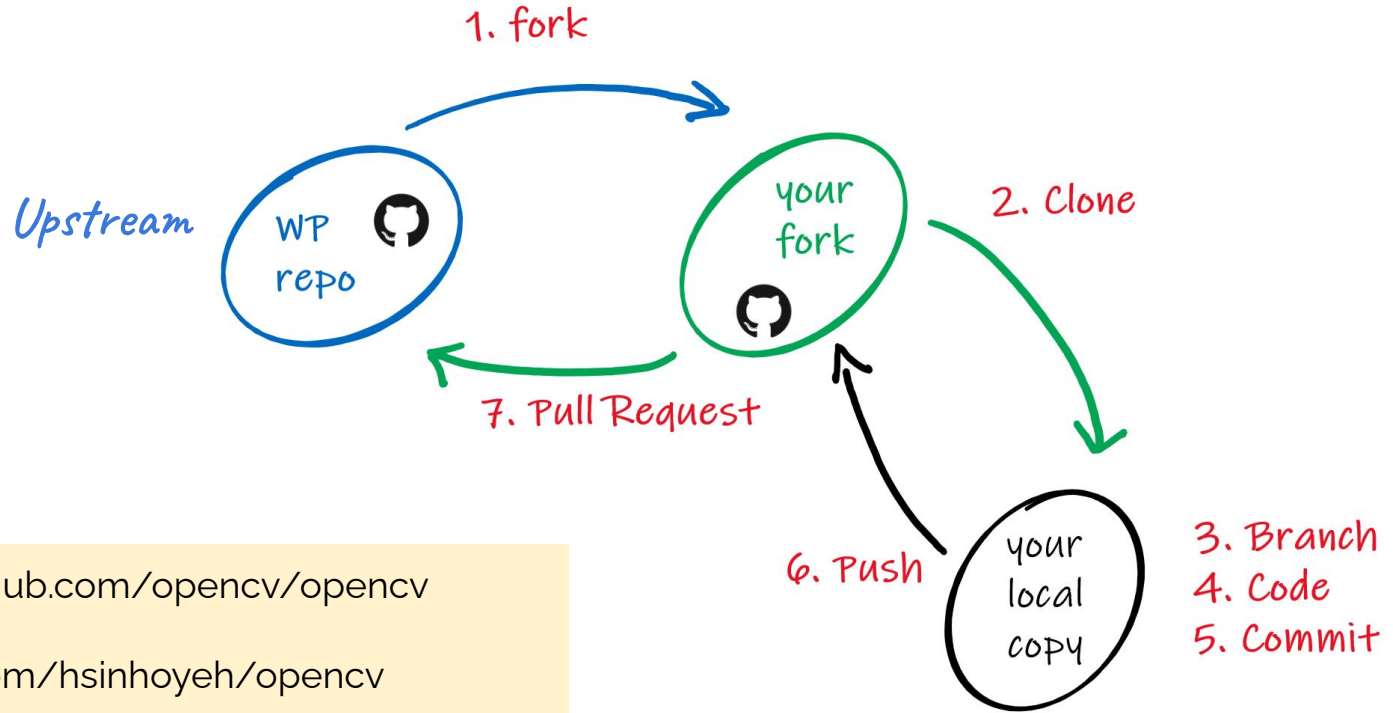
#5544 opened 5 days ago by amitza

1

Git Basic



Pull-Request flow in Github

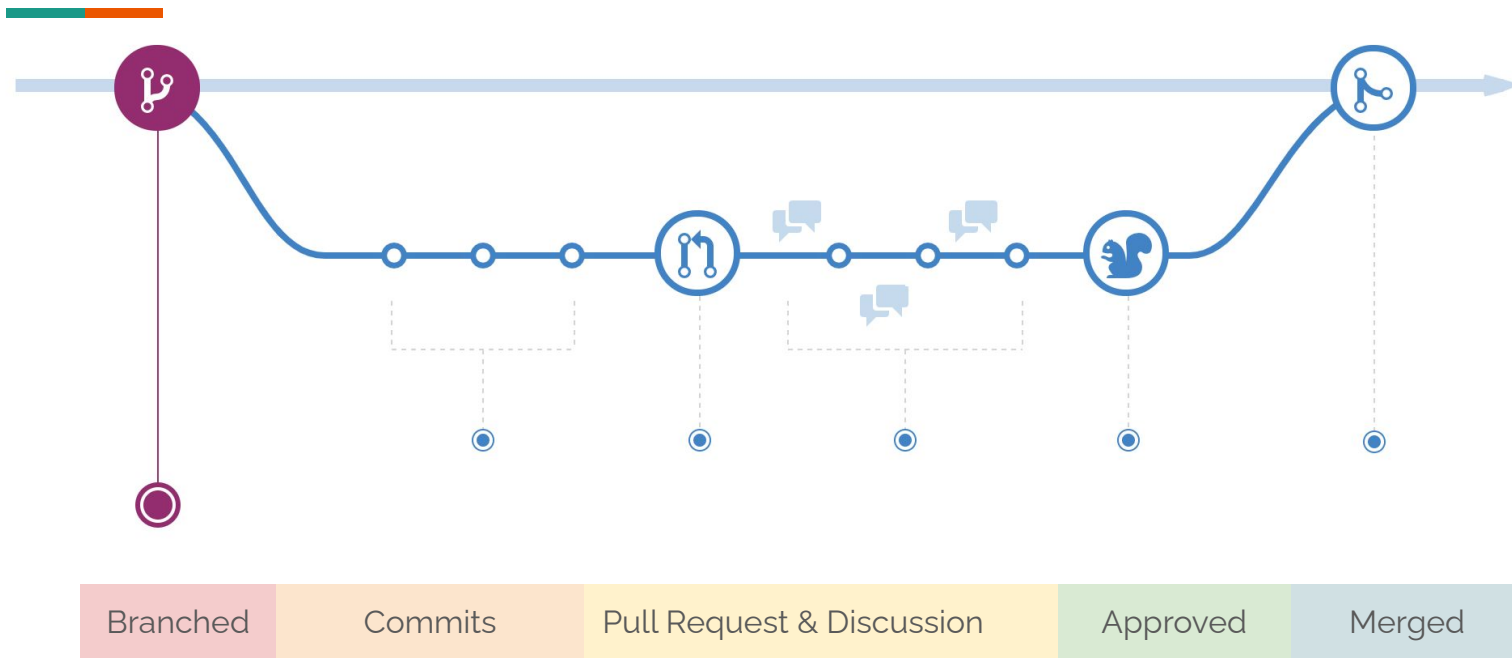


Upstream: github.com/opencv/opencv

Fork: github.com/hsinhoyeh/opencv

Ref: <https://developer.wordpress.org/block-editor/contributors/code/git-workflow/>

Github Flow: branch-based workflow



81 Open ✓ 3,342 Closed

Author ▾ Label ▾ Projects ▾ Milestones ▾ Reviews ▾ Assignee ▾ Sort ▾

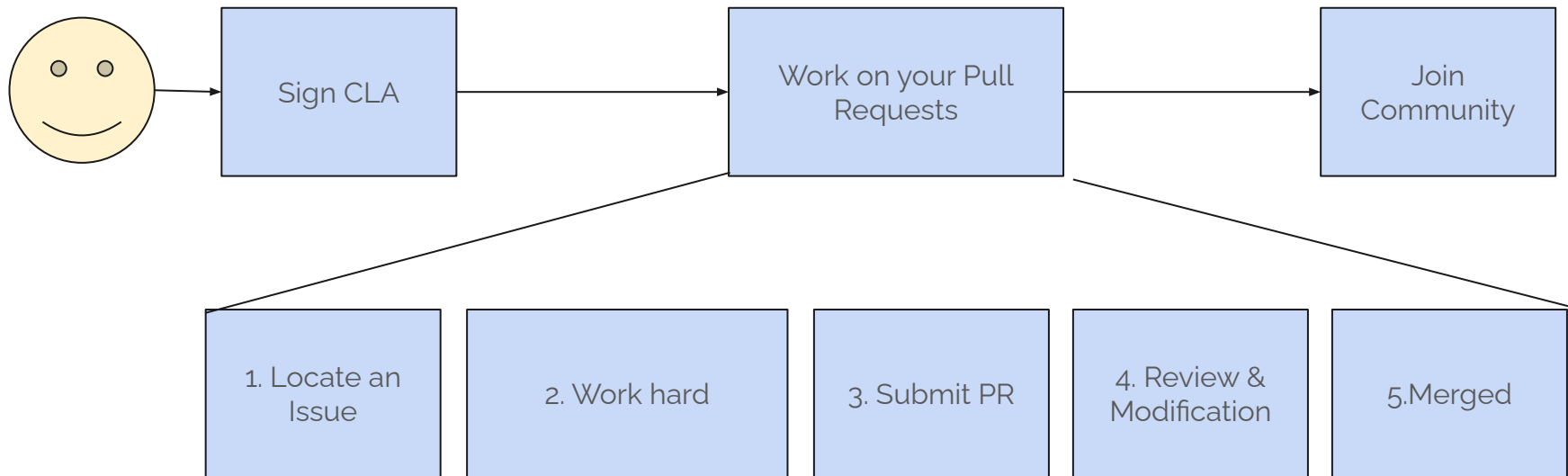
- chore: [Research] Add toxicity detection pipeline • cla: yes needs-ok-to-test size/S #5573 opened 1 hour ago by TheDutchDevil 1 of 1 4
- feat(sdk.v2): reimplement custom job support • cla: yes size/XL** #5572 opened 4 hours ago by chensun 1 of 1 1
- Add OWNERS for v2 folder • cla: yes lgtn size/XS #5571 opened 10 hours ago by Bobgy 0 of 1 2
- Add zijianjoy to frontend/OWNERS • approved cla: yes size/XS #5570 opened 10 hours ago 2
- chore(deps): bu #5569 opened 11 hours ago
- chore(google_c #5554 opened 3 days ago size/M
- feat(sdk): Add a #5552 opened 3 days ago
- chore(v2): fix in #5543 opened 5 days ago size/M Ref: <https://www.conventionalcommits.org/en/v1.0.0/>
- chore(samples): v2 compatible exit handler sample • cla: yes size/L #5544 opened 5 days ago 4

Commit Message Format:

`<type>(<scope>): <short summary>`

- Summary in present tense. Not capitalized. No period at the end.
- Scope: common|compiler|authentication|core|
- Type: chore, docs, feat, fix, refactor, style, or test.

How to be a Contributor?



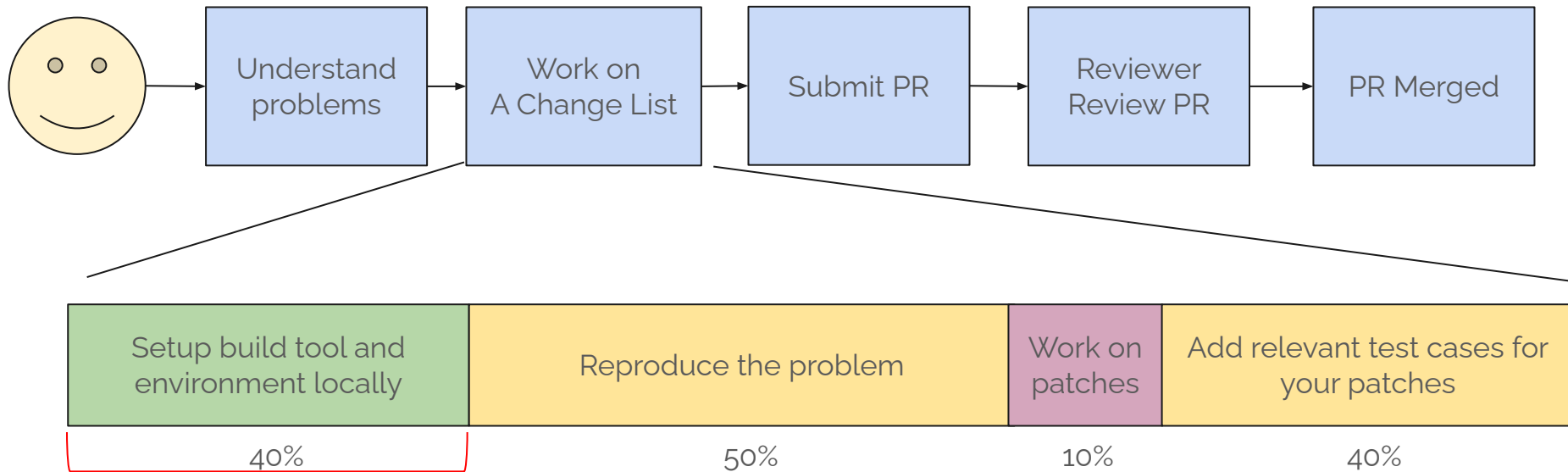
Ref: <https://www.kubeflow.org/docs/about/contributing/>



Find your first issue

1. Existing issues
 - a. Issues labelled with `good first issue`: [opencv/opencv](#)
 - b. Issues labelled with `help wanted`: [opencv/opencv](#)
2. Organic issues
 - a. Create an issue to describe issues you encountered and also provide a solution (a.k.a Pull-request) for it

Why is the initial contribution significant?



One-time Work



OpenCV Case Study

opencv/opencv#24424

github.com/opencv/opencv/issues/24424

opencv / opencv

<> Code Issues 2.4k Pull requests 118 Actions Wiki Security Insights

Update DNN Optical Flow sample with RAFT model #24424

Open dkurt opened this issue 2 weeks ago · 2 comments

dkurt commented 2 weeks ago

Describe the feature and motivation

Originally, sample https://github.com/opencv/opencv/blob/4.x/samples/dnn/optical_flow.py created for FlowNet v2. Recently, @Aser-Abdelfatah and @fengyuantau enabled support of RAFT model. Sample should work with both models.

Additional context

No response

4

dkurt added feature category: sample good first issue category: dnn labels 2 weeks ago

parthasdey2304 commented last week

would you please assign this to me

Assignees

No one assigned

Labels

category: dnn category: samples feature good first issue

Milestone

No milestone

Development

Successfully merging a pull request may close this issue.

Update optical_flow.py officialhemant511/opencv

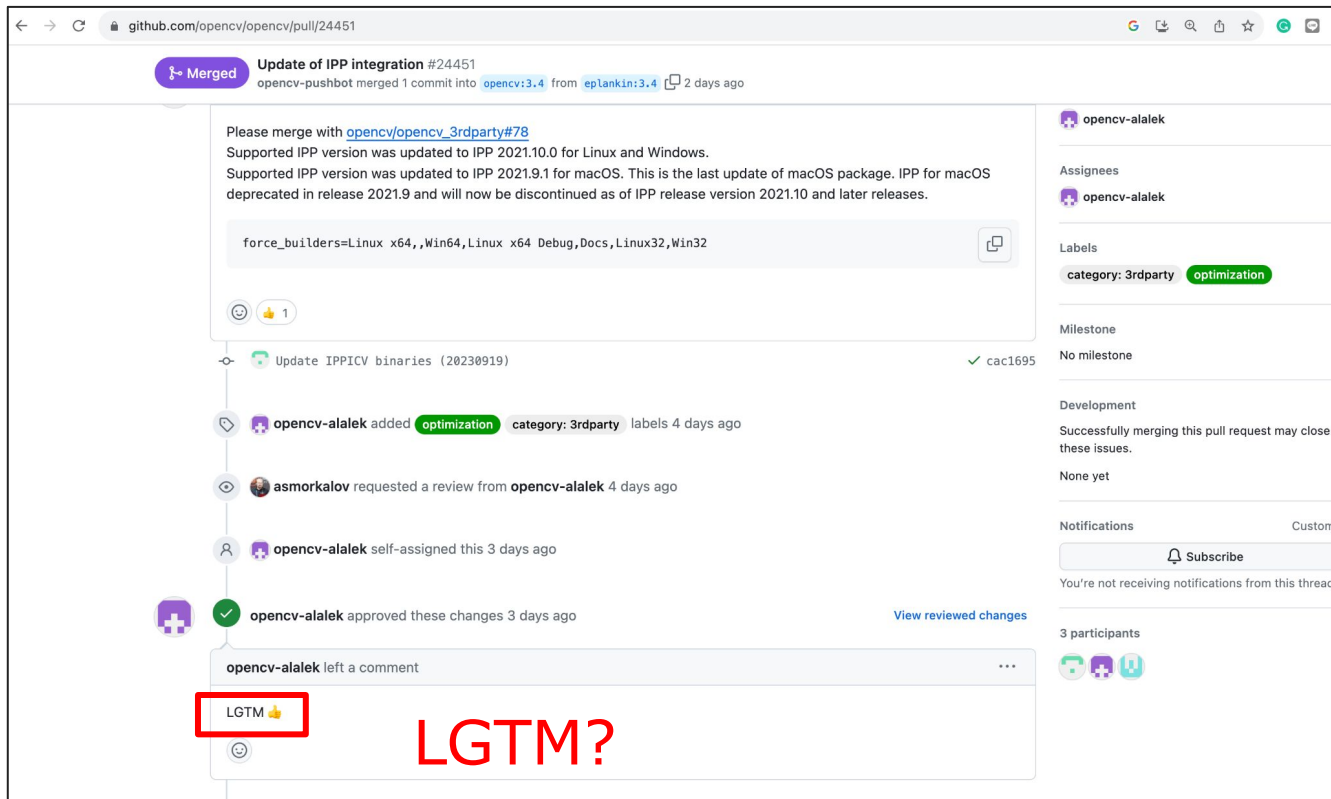
Notifications

Subscribe

You're not receiving notifications from this thread.

3 participants

opencv/opencv#24451



← → ↻ github.com/opencv/opencv/pull/24451

Merged Update of IPP integration #24451
opencv-pushbot merged 1 commit into [opencv:3.4](#) from [eplankin:3.4](#) 2 days ago

Please merge with [opencv/opencv_3rdparty#78](#)
Supported IPP version was updated to IPP 2021.10.0 for Linux and Windows.
Supported IPP version was updated to IPP 2021.9.1 for macOS. This is the last update of macOS package. IPP for macOS deprecated in release 2021.9 and will now be discontinued as of IPP release version 2021.10 and later releases.

```
force_builders=Linux x64,,Win64, Linux x64 Debug, Docs, Linux32, Win32
```

Update IPPICV binaries (20230919) ✓ cac1695

opencv-alalek added **optimization** category: 3rdparty labels 4 days ago

asmorkalov requested a review from opencv-alalek 4 days ago

opencv-alalek self-assigned this 3 days ago

opencv-alalek approved these changes 3 days ago [View reviewed changes](#)

opencv-alalek left a comment

LGTM 🍏

LGTM?

opencv-alalek

Assignees

opencv-alalek

Labels

category: 3rdparty **optimization**

Milestone

No milestone

Development

Successfully merging this pull request may close these issues.

None yet

Notifications [Custom](#)

[Subscribe](#)

You're not receiving notifications from this thread

3 participants

opencv/opencv#24446

github.com/opencv/opencv/pull/24446

Closed Update optical_flow.py #24446
officialhemant511 wants to merge 1 commit into opencv:4.x from officialhemant511:patch-1

```
... @@ -59,17 +47,17 @@ def motion_to_color(self, flow):
59 47
60 48
61 49 if __name__ == '__main__':
62 - parser = argparse.ArgumentParser(description='Use this script to calculate optical flow us
50 + parser = argparse.ArgumentParser(description='Use this script to calculate optical flow us
```

dkurt 5 days ago Member ...
Sample should keep work with FlowNetv2
😊

officialhemant511 4 days ago Author ...
yeap.....i think it will work well...
😊

Reply...

mshabunin commented 4 days ago
Looks like this is garbage PR. 🍌 1

mshabunin closed this 4 days ago

Looks like this is garbage PR.



Other Case Study

nats-io/nats.go#22 Case Study

Closed

goroutine leaking when calls BindRecvChan #22
hsinhoyeh opened this issue on Jul 26, 2014 · 1 comment

in the underlying implementation, nats defines a callback (i.e. cb) and then call the Subscribe from a conn
<https://github.com/apcera/nats/blob/master/netchan.go#L81>

```
_, err := c.Conn.subscribe(subject, queue, cb)
```

in the conn.subscribe, nats spawns a go routine to send the message:
<https://github.com/apcera/nats/blob/master/nats.go#L1198>

```
go nc.deliverMsgs(sub.mch)
```

and that go routine will be terminated "only if sub.mch is closed", which is called via conn.Unsubscribe().
<https://github.com/apcera/nats/blob/master/nats.go#L1244>

```
func (nc *Conn) unsubscribe(sub *Subscription, max int) error {
```

The bug actually happened, when we look BindRecvChan again: even the caller has closed the channel, the go routine who serves deliverMsg will never be destroyed. We actually have an additional zombie go routine when we call BindRecvChan every time.

Projects

None yet

Milestone

No milestone

Linked pull requests

Successfully merging a pull request may close this issue.

None yet

Notifications

Customize

Unsubscribe

You're receiving notifications because you were assigned.

2 participants

hsinhoyeh referenced this issue in hsinhoyeh/nats on Jul 26, 2014

workaround the goroutine leak problem in apcera/nats/netchan.go 0d117eb

derecollison added the bug label on Jul 26, 2014

derecollison self-assigned this on Jul 26, 2014

derecollison commented on Jul 26, 2014

The proposed solution could affect ordering of messages due to the read. But great bug report. Will fix. Thanks.

googleapis/google-cloud-go#170 Case Study (1/2)

Closed

[bigtable] bttest needs UpdateColumnFamily support #170

hsinhoyeh opened this issue on Sep 15, 2015 · 3 comments

googleapis/google-cloud-go#170

```
diff --git a/google.golang.org/cloud/bigtable/bttest/inmem.go b/google.golang.org/cloud/bigtable/bttest/inmem.go
index 5c4bc09..f4dbe03 100644
--- a/google.golang.org/cloud/bigtable/bttest/inmem.go
+++ b/google.golang.org/cloud/bigtable/bttest/inmem.go
+func (s *server) UpdateColumnFamily(ctx context.Context, req *bttdpb.ColumnFamily) (*bttdpb.ColumnFamily, error) {
+    return nil, errors.New("no op")
+}
```

we need it because we call SetGCPolicy to configure the maximum versions stored in a cell

```
adminClient.SetGCPolicy( ctx, tableName, columnFamily, bgt.MaxVersionsPolicy(1))
```

and in the underlying implementation, it will first call GetTable and then update its column family.

```
func (ac *AdminClient) SetGCPolicy(ctx context.Context, table, family string, policy GCPolicy) error {
    prefix := ac.clusterPrefix()
    tbl, err := ac.tClient.GetTable(ctx, &bttspb.GetTableRequest{
        Name: prefix + "/tables/" + table,
    })
    if err != nil {
        return err
    }
    fam, ok := tbl.ColumnFamilies[family]
    if !ok {
        return fmt.Errorf("unknown column family %q", family)
    }
    fam.GcRule = policy.proto()
    _, err = ac.tClient.UpdateColumnFamily(ctx, fam)
    return err
}
```

<https://github.com/GoogleCloudPlatform/gcloud-golang/blob/master/bigtable/admin.go#L161>

api: bigtable

triage me

Projects

None yet

Milestone

No milestone

Linked pull requests

Successfully merging a pull request may close this issue.

None yet





Notifications


Customize

Unsubscribe

You're receiving notifications because you authored the thread.

4 participants





hsinhoyeh commented on Sep 16, 2015

Contributor Author

hi,

googleapis/google-cloud-go#170 Case Study (2/2)

Gerrit CHANGES DOCUMENTATION BROWSE

Merged 3471 bigtable/bttest: support GetTable/UpdateColumnFamily

Change Info [SHOW ALL](#)

Submitted Sep 23, 2015

Owner [Hsin-Ho Yeh](#)

Uploader [David Symonds](#)

Committer [Gerrit Code Review](#)

Reviewers [David Symonds](#)

Repo | Branch [gocloud](#) | [master](#)

Topic [✎](#)

Submit requirements

✓ Code-Review +2 [David Symonds](#)

bigtable/bttest: support GetTable/UpdateColumnFamily

Fixes [#166](#).

Fixes [#178](#).

Change-Id: [Ic2b3698583923842253d591b137c7192685ecef0](#)

Comments [25 resolved](#)

Files	Comments	Findings
Base → Patchset 4 bcca3e9	DOWNLOAD INCLUDED IN EXPAND ALL	
File	Comments	Size
Commit message		Delta
bigtable/bttest/inmem.go		+79 -9
		+79 -9

[Change Log](#)

[EXPAND ALL](#)

Hsin-Ho Yeh	Uploaded patch set 1.	Patchset 1 Sept 19, 2015 16:32
David Symonds	16 comments Thanks for doing this. Please check the CLA details first. If you're working for a company then you might need to get them to approve your contribution.	Patchset 1 Sept 22, 2015 13:10
Hsin-Ho Yeh	Uploaded patch set 2.	Patchset 2 Sept 22, 2015 23:05
Hsin-Ho Yeh	16 comments the updated patch address the comments.	Patchset 2 Sept 22, 2015 23:19
David Symonds	10 comments Nearly there.	Patchset 2 Sept 23, 2015 07:12

istio/istio#46327 Case Study

github.com/istio/istio/issues/46327

Closed 3 of 16 tasks Unable to use Filesystem HTTP Cache EnvoyFilter #46327
Stono opened this issue on Aug 5 · 7 comments · Fixed by istio/proxy#5008

one interesting finding is that when I run `istio/proxyv2:1.17.6` (envoy version is v1.25.9-dev) the envoy does not have `file_system_http_cache` plugin

```
docker run --it --entrypoint envoy docker.io/istio/proxyv2:1.17.6 | grep envoy.http.cache
[2823-10-10 04:26:37.077] [I] [info] [main] [external/envoy/source/server/server.cc:408] envoy.http.cache: ....0}
```

but when we got [the binary](#) from envoy, the plugin is there.

```
./envoy-1.25.9-linux-x86_64 2>61 | grep envoy.http.cache
[2823-10-10 04:27:52.579] [792396] [info] [main] [source/server/server.cc:408] envoy.http.cache: envoy.ext...
```

hainhoyeh commented 3 weeks ago · edited

@zirain is it possible that the cache filter is not included in [here](#)?

```
# CacheFilter plugins
#
"envoy.extensions.http.cache.simple": ":///source/extensions/http/cache/simple_http_cache:conf:
```

Here is [extensions build configs](#) from envoyproxy/envoy for referencing:

```
#
# CacheFilter plugins
#
"envoy.extensions.http.cache.file_system_http_cache": ":///source/extensions/http/cache/file_system_http_cache:conf:"
"envoy.extensions.http.cache.simple": ":///source/extensions/http/cache/simple_http_cache:conf:"
```

zirain commented 3 weeks ago Member

you got it, let me fix it.

天下武功，唯快不破

torvalds/linux#661 Case Study - Not all CLs are valuable

Fixed typo: extra space. #661
jacob-macleod wants to merge 1 commit into `torvalds:master` from `jacob-macleod:master`

jacob-macleod commented on Mar 8, 2019

There was a double space.

10 2 13 1 6

Fixed typo: extra space. Verified ee9ca9c

KernelPRBot commented on Mar 8, 2019

Hi @jacob-and-nathan!

Thanks for your contribution to the Linux kernel!

Linux kernel development happens on mailing lists, rather than on GitHub - this GitHub repository is a read-only mirror that isn't used for accepting contributions. So that your change can become part of Linux, please email it to us as a patch.

Sending patches isn't quite as simple as sending a pull request, but fortunately it is a well documented process.

Here's what to do:

- Format your contribution according to kernel requirements
- Decide who to send your contribution to
- Set up your system to send your contribution as an email
- Send your contribution and wait for feedback

How do I format my contribution?

The Linux kernel community is notoriously picky about how contributions are formatted and sent. Fortunately, they have documented their expectations.

Firstly, all contributions need to be formatted as patches. A patch is a plain text document showing the change you want to make to the code, and documenting why it is a good idea.

You can create patches with `git format-patch`.

Secondly, patches need 'commit messages' which is the human-friendly documentation explaining what the change is.

Assignees
No one assigned

Labels
None yet

Projects
None yet

Milestone
No milestone

Notifications [Customize](#)
[Subscribe](#)
You're not receiving notifications from this thread.

5 participants



Conclusion

- Open source is a public artifacts that help you grow a reputation and a career.
- The work to contribute to Open Source could take days or weeks, but the rewards could keep years.
- Let's make changes to the software world now, even steps are small :)

An aerial photograph of the New York City skyline at dusk. The sky is a mix of dark blue and orange, with scattered clouds. The city is densely packed with skyscrapers, many of which are illuminated with their lights. The Empire State Building is prominent in the center, with its top lit in red and green. The Hudson River is visible in the background, with the New York-New York Hotel & Casino's replica of the Empire State Building on the right side of the image.

Thank You
Questions?

Contact Us:
partnership@footprint-ai.com