My {cdcfluview} package started tossing erros on CRAN just over a week ago when the CDC added an extra parameter to one of the hidden API endpoints that the package wraps. After a fairly hectic set of days since said NOTE came, I had time this morning to poke at a fix. There are *alot* of tests, so after successful debugging session I was awaiting CRAN checks on various remotes as well as README builds and figured I’d keep up some practice with another, nascent, package of mine, {swiftr}, which makes it dead simple to build R functions from Swift code, in similar fashion to what Rcpp::cppFunction() does for C/C++ code.

macOS comes with a full set of machine learning/AI libraries/frameworks that definitely have “batteries included” (i.e. you can almost just make one function call to get 90-95% what you want without even training new models). One of which is text extraction from Apple’s computer Vision framework. I thought it’d be a fun and quick “wait mode” distraction to wrap the VNRecognizeTextRequest() function and use it from R.

To show how capable the default model is, I pulled a semi-complex random image from DDG’s image search:



Let’s build the function (you need to be on macOS for this; exposition inine):

library(swiftr)

swift\_function( code = '

import Foundation import CoreImage import Cocoa import Vision

@\_cdecl ("detect\_text")

public func detect\_text(path: SEXP) -> SEXP {

// turn R string into Swift String so we can use it

let fileName = String(cString: R\_CHAR(STRING\_ELT(path, 0))) var res: String = ""

var out: SEXP = R\_NilValue

// get image into the right format

if let ciImage = CIImage(contentsOf: URL(fileURLWithPath:fileName)) {

let context = CIContext(options: nil)

if let img = context.createCGImage(ciImage, from: ciImage.extent) {

// setup comptuer vision request

let requestHandler = VNImageRequestHandler(cgImage: img)

// start recognition

let request = VNRecognizeTextRequest() do {

try requestHandler.perform([request])

// if we have results

if let observations = request.results as? [VNRecognizedTextObservation] {

// paste them together

let recognizedStrings = observations.compactMap { observation

in

observation.topCandidates(1).first?.string

}

res = recognizedStrings.joined(separator: "\\n")

}

} catch { debugPrint("\\(error)")

}

}

}

res.withCString { cstr in out = Rf\_mkString(cstr) }

return(out)

} ')

The detect\_text() is now available in R, so let’s see how it performs on that image of signs:

detect\_text(path.expand("~/Data/signs.jpeg")) %>% stringi::stri\_split\_lines() %>%

unlist()

## [1] "BEWILDERED" "UNCLEAR" "nAZEU" "UNCERTAIN" "VISA" "INSURE"

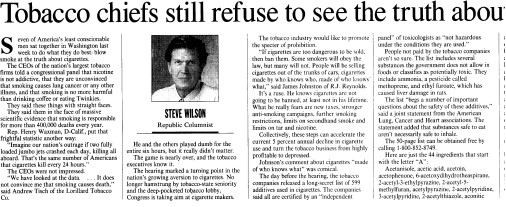
## [7] "ATED" "MUDDLED" "LOsT" "DISTRACTED" "PERPLEXED" "CONFUSED"

## [13] "PUZZLED"

It works super-fast and gets far more correct than I would have expected.

Toy examples aside, it also works pretty well (as one would expect) on “real” text images, such

as this example from the Tesseract test suite:



detect\_text(path.expand("~/Data/tesseract/news.3B/0/8200\_006.3B.tif"))

%>%

stringi::stri\_split\_lines() %>%

|  |  |  |  |
| --- | --- | --- | --- |
| unlist() | | | |
| ## | [1] | "Tobacco chiefs still refuse to see the truth abou" |  |
| ## | [2] | "even of America's least conscionable" |  |
| ## | [3] | "The tobacco industry would like to promote" |  |
| ## | [4] | "men sat together in Washington last" |  |
| ## | [5] | "under the conditions they are used.'" |  |
| ## | [6] | "week to do what they do best: blow" |  |
| ## | [7] | "the specter of prohibition." |  |
| ## | [8] | "panel\" of toxicologists as \"not hazardous" |  |
| ## | [9] | "smoke at the truth about cigarettes." |  |
| ## | [10] | "'If cigarettes are too dangerous to be sold," |  |
| ## | [11] | "then ban them. Some smokers will obey the" |  |
| ## | [12] | "People not paid by the tobacco companies" |  |
| ## | [13] | "aren't so sure. The list includes several" |  |
| ## | [14] | "The CEOs of the nation's largest tobacco" |  |
| ## | [15] | "firms told congressional panel that nicotine" |  |
| ## | [16] | "law, but many will not. People will be selling" |  |
| ## | [17] | "iS not addictive, that they are unconvinced" |  |
| ## | [18] | "cigarettes out of the trunks of cars, cigarettes" |  |
| ## | [19] | "substances the government does not allow in" |  |
| ## | [20] | "foods or classifies as potentially toxic. They" |  |
| ## | [21] | "that smoking causes lung cancer or any other" |  |
| ## | [22] | "made by who knows who, made of who knows include ammonia, | a |
| pesticide called" | | | |
| ## | [23] | "illness, and that smoking is no more harmful" | |
| ## | [24] | "what,\" said James Johnston of R.J. Reynolds." | |
| ## | [25] | "than drinking coffee or eating Twinkies." | |
| ## | [26] | "It's a ruse. He knows cigarettes are not" | |
| ## | [27] | "methoprene, and ethyl furoate, which has" | |
| ## | [28] | "They said these things with straight taces." | |
| ## | [29] | "going to be banned, at leasi not in his lifetime." | |
| ## | [30] | "caused liver damage in rats." | |
| ## | [31] | "The list \"begs a number of important" | |
| ## | [32] | "They said them in the face of massive" | |
| ## | [33] | "STEVE WILSON" | |
| ## | [34] | "What he really fears are new taxes, stronger" | |
| ## | [35] | "questions about the safety of these additives,\"" | |

|  |  |  |  |
| --- | --- | --- | --- |
| ## | [36] | "scientific evidence that smoking is responsible" |  |
| ## | [37] | "anti-smoking campaigns, further smoking" |  |
| ## | [38] | "said a joint statement from the American" |  |
| ## | [39] | "for more than 400,000 deaths every year." |  |
| ## | [40] | "restrictions, limits on secondhand smoke and" |  |
| ## | [41] | "Rep. Henry Waxman, D-Calif., put that" |  |
| ## | [42] | "Republic Columnist" |  |
| ## | [43] | "Lung, Cancer and Heart associations. The" |  |
| ## | [44] | "limits on tar and nicotine." |  |
| ## | [45] | "statement added that substances safe to eat" |  |
| ## | [46] | "frightful statistic another way:" |  |
| ## | [47] | "Collectively, these steps can accelerate the" |  |
| ## | [48] | "\"Imagine our nation's outrage if two fully" |  |
| ## | [49] | "He and the others played dumb for the" |  |
| ## | [50] | "current 5 percent annual decline in cigarette" |  |
| ## | [51] | "aren't necessarily safe to inhale." |  |
| ## | [52] | "The 50-page list can be obtained free by" |  |
| ## | [53] | "loaded jumbo jets crashed each day, killing all" |  |
| ## | [54] | "entire six hours, but really didn't matter." |  |
| ## | [55] | "use and turn the tobacco business from highly" |  |
| ## | [56] | "calling 1-800-852-8749." |  |
| ## | [57] | "aboard. That's the same number of Americans" |  |
| ## | [58] | "The game i nearly over, and the tobacco" |  |
| ## | [59] | "profitable to depressed." |  |
| ## | [60] | "Johnson's comment about cigarettes \"made" |  |
| ## | [61] | "Here are just the 44 ingredients that start" |  |
| ## | [62] | "that cigarettes kill every 24 hours.'" |  |
| ## | [63] | "executives know it." |  |
| ## | [64] | "with the letter \"A\":" |  |
| ## | [65] | "The CEOs were not impressed." |  |
| ## | [66] | "The hearing marked a turning point in the" |  |
| ## | [67] | "of who knows what\" was comical." |  |
| ## | [68] | "Acetanisole, acetic acid, acetoin," |  |
| ## | [69] | "\"We have looked at the data." |  |
| ## | [70] | "It does" |  |
| ## | [71] | "nation's growing aversion to cigarettes. No" |  |
| ## | [72] | "The day before the hearing, the tobacco" |  |
| ## | [73] | "acetophenone,6-acetoxydihydrotheaspirane," |  |
| ## | [74] | "not convince me that smoking causes death,\"" |  |
| ## | [75] | "2-acetyl-3-ethylpyrazine, 2-acetyl-5-" |  |
| ## | [76] | "said Andrew Tisch of the Lorillard Tobacco" |  |
| ## | [77] | "longer hamstrung by tobacco-state seniority" |  |
| ## | [78] | "companies released a long-secret list of 599" |  |
| ## | [79] | "Co." |  |
| ## | [80] | "and the deep-pocketed tobacco lobby," |  |
| ## | [81] | "methylfuran, acetylpyrazine, 2-acetylpyridine," |  |
| ## | [82] | "Congress is taking aim at cigarette makers." |  |
| ## | [83] | "additives used in cigarettes. The companies" |  |
| ## | [84] | "said all are certified by an \"independent" |  |
| ## | [85] | "3-acetylpyridine, 2-acetylthiazole, aconitic" | … |