subvision

vision-driven subtitle detection

Extract. Review. Export.

SubVision User Guide

App 1.2 · Build 6 · Manual 2.2

Contents

1. Getting started 3

2. Interface and opening a video 4

3. Define the OCR region 5

4. Recognition settings and languages 6

5. Detection and image enhancement 7

6. Scan, Append Scan and scan range 8

7. Set and clear IN and OUT 10

8. Review, search and warnings 11

9. Edit captions and re-read 12

10. Export SRT and leave safely 13

11. App Store and Privacy 14

12. Troubleshooting and diagnostics 15

13. Keyboard shortcuts and Help 16

This guide describes the current version of SubVision. Help and PDF share one source. Updates replace the relevant instructions rather than appending release notes.

1. Getting started

SubVision extracts burned-in subtitles from video on your Mac using optical character recognition (OCR). You can review the detected captions, correct their text and timing, and export an SRT file. Recognition works on the visible video image; it does not extract an existing embedded subtitle track.

This guide describes SubVision 1.2, build 6. Help and the downloadable guide contain the same instructions. Manual revision 2.0 replaces the earlier guide rather than adding release notes to it.

Your first extraction

1. Choose File > Open Video, or click Open Video in the toolbar. You can also drop a local video into the window.

2. Find a frame that contains subtitles and position the yellow Region of Interest (ROI) rectangle around them.

3. Open Settings in the toolbar. Choose the recognition mode and, when known, the primary language.

4. Optionally set scan IN and OUT to test a short section. Otherwise Scan examines the full video.

5. Click Scan. The central progress panel shows the current timecode, percentage and analyzed frame count.

6. Review the caption list, fix text or timing, and use Re-read Selection for a difficult passage.

7. Click Export SRT and choose a destination. Export includes the complete caption list, including captions hidden by Search or Warnings only.

OCR is sensitive to font size, background contrast, compression and motion. Always review the results before publishing or using the SRT in an edit.

Supported files and privacy

Use local video files on your Mac or accessible connected storage. MP4, M4V, MOV and QT are common choices. The file picker accepts video types, while loading checks an extension allowlist, AVFoundation playability and the presence of a video track. Additional accepted extensions include AVI, MPEG, MPG, M2V, MTS, M2TS, TS, DV, 3GP and 3G2; acceptance does not guarantee that the codec can be decoded. MKV is not in the current allowlist. An invalid video leaves the existing session intact.

SubVision targets macOS 14 or later. OCR runs locally through Apple's Vision framework; this extraction workflow does not upload your video or captions. Access to user-selected files is managed by the macOS sandbox. Keep the source available during scanning, including files on external drives.

2. Interface and opening a video

The main window contains a resizable caption sidebar and the video area. The toolbar includes Settings, Scan, Append Scan and Open Video. Settings remains in the toolbar before and after a video is loaded. Reset ROI is available when a video is loaded and scanning is idle.

Sidebar and Settings

Drag the sidebar divider to adjust its width. Use the sidebar toolbar control to show or hide the sidebar. Settings opens Preferences within the sidebar; its controls scroll vertically so that all options remain accessible at different window sizes. Done closes Preferences. Reset All restores the standard detection and image-filter settings, including the default recognition configuration.

Search appears below the logo in the caption sidebar. When at least one caption has a warning, Warnings only appears nearby. Caption controls at the bottom let you adjust timing, re-read, export and clear results.

Video controls and Status

Use the video player for playback and seeking. The bottom bar provides one-second jumps, frame steps, the current timecode and buttons for scan IN and OUT. The main window title shows the file name and available duration, dimensions and frame rate.

The Status overlay displays the recognition group, primary language if selected, detection preset and visual preset. Language optimization appears when Language correction is enabled. With scan points set, Status also displays the scan range; set timecodes use the app's highlight color. Status can be moved by dragging it.

Progress, scan results and diagnostics are centered on the entire main-window content area, rather than the video area or sidebar. Confirmation and file dialogs are attached to the main window and use the same centering policy.

Replace the current video safely

Open Video in the toolbar and File > Open Video perform the same operation. Opening another video starts a new in-memory session: captions, scan points and scan coverage are reset. ROI and recognition/image preferences are retained.

If the current captions have not been exported, or have changed since export, SubVision asks before opening the file picker. Choose Export SRT to save first, Discard and Open to continue without saving, or Cancel to keep working. A successful export is required to continue through Export SRT. Canceling the export or a write failure keeps the current session and prevents the video picker from opening.

Dropping a replacement video also uses the protection. The replacement is validated before it is loaded; a rejected file does not remove your work. After an export, subsequent text or timing changes require a new export.

3. Define the OCR region

The yellow rectangle is the Region of Interest. Only the video image within that region is passed to OCR. A tight, well-positioned ROI is usually the most useful improvement you can make.

Move and resize

Find a frame with a representative subtitle. Drag inside the rectangle to move it. Drag a corner or an edge handle to resize it; all eight handles can be used. The rectangle stays within the displayed video and enforces a minimum size. While resizing one side, the opposite side remains fixed.

Position the rectangle around all subtitle lines, including descenders and outlines. Avoid channel logos, watermarks, credits and unrelated graphics. If subtitle positions change within a video, process those sections separately with Append Scan and a suitable ROI for each section.

The cropped preview is refreshed when the adjustment is released. Preview work is suspended during manipulation to improve responsiveness; when paused, the last video frame can be used to refresh the final crop. Rotated video is oriented consistently for preview and batch recognition.

Reset and check the crop

Click Reset ROI to restore the default region. Check the preview on both short and long captions and on bright and dark scenes before committing to a long scan. Image enhancement is applied to the cropped recognition image in both batch scanning and re-reading; it does not alter the original video file.

4. Recognition settings and languages

Open Settings and choose Text Recognition Mode. Primary-language choices come from the accurate Vision models available on this Mac and can vary with macOS.

Recognition groups

Mode	Intended content
Latin (Western)	Supported languages using Latin script, including English and Italian when their models are available.
Cyrillic	Supported Cyrillic-script languages, such as Russian and Ukrainian.
Asian (CJK)	Supported Chinese, Japanese and Korean models, including simplified/traditional Chinese and Cantonese models when available.
Arabic	Supported Arabic-script language models.
Universal (Auto)	Automatic selection among the languages supported by Vision, including scripts outside the named groups.

Groups guide model selection; they do not strictly filter output characters. Use Universal for an unknown language, or a specific group when you know the content.

Primary language

Primary language gives one model first priority while retaining the other group languages. Group default uses the reported order. For English content, choose Latin and the English model offered by the menu.

An incompatible group clears the primary language. Universal uses automatic detection. Detection presets preserve group, primary language and correction.

Language correction and optimization

Language correction is enabled by default. It uses language dictionaries to help recognize ordinary words, but can change proper names, abbreviations or unusual spellings. Disable it when those changes are undesirable and compare a short test scan. Status shows Language optimization while this option is enabled; it refers to the same dictionary-correction setting.

Valid short words are not discarded merely because they contain one or two letters. Confidence thresholds and repeated-frame confirmation still apply. Punctuation-only detections are filtered out.

5. Detection and image enhancement

Standard, Precision and Hardcore are starting points for detection timing and confidence. Selecting one changes the advanced detection parameters while retaining your language choices. Editing a parameter can make the preset Custom. Changes made during a running scan affect a subsequent scan; the current batch uses its captured configuration.

Detection parameters

Control	Effect
Sample Rate	Interval between analyzed frames. A smaller displayed millisecond value samples more often and can catch briefer captions at increased scan cost.
Min Confidence	Minimum certainty required to accept a recognized line. Raising it reduces uncertain text but may omit faint captions.
Confirm Frames	Number of matching sampled frames required to confirm a caption. Higher values reduce flicker but can miss very brief text.
Gap Threshold	Joins matching text separated by a small gap within a batch. It does not restore text that OCR failed to detect.
Export Offset	Global timing shift applied only when writing SRT. It does not shift the displayed list timecodes.

For brief subtitles, first test a shorter sample interval and fewer confirmation frames. For unstable background text, tighten the ROI before raising thresholds. Very strict settings can produce an empty result even when text is visible.

Visual enhancement

Original, Enhanced and High Contrast provide image-filter presets. You can also adjust Contrast, Brightness and Sharpen, enable Grayscale or Invert, and enable B/W Threshold with its Threshold Level.

Use contrast or grayscale to simplify a busy image. Try Invert for a difficult light-on-light passage. Threshold converts the crop to black and white; an unsuitable level can remove thin strokes, so inspect the preview. Filters affect the recognition crop, not playback media or the source file.

6. Scan, Append Scan and scan range

Scan and Append Scan are distinct operations. Both use the current ROI, recognition settings and image filters. While processing, these buttons remain visible but disabled. Progress is shown only in the central panel; Stop is available there and in the toolbar.

Scan creates a replacement list

Scan analyzes IN to OUT, defaulting to the full video when neither is set, and replaces the existing caption list with the results of that scan. If captions already exist, it asks for confirmation before replacing them. Export SRT preserves a copy first; Discard and Scan proceeds without keeping the old list; Cancel makes no change.

Use Scan for an initial extraction or when deliberately rebuilding the list. If IN/OUT describes only a short test range, Scan still creates a replacement list for that range. Use Append Scan when you need to keep captions elsewhere in the video.

Append Scan preserves the rest

Append Scan is available whenever a video is loaded and scanning is idle. It scans the chosen interval and replaces caption portions inside that interval, while preserving portions outside it. Repeating a scan of the same interval updates those results rather than adding duplicate copies.

The starting point follows this order:

1. An explicit scan IN, if set.

2. The saved continuation point from an interrupted or failed scan.

3. The end of the last recorded scanned interval.

4. The beginning of the video.

OUT sets the end; without OUT, the end is the video's duration. SubVision records scanned intervals even when they contain no captions, so continuation does not depend on the last detected caption. If no video remains within the chosen range, set a new IN/OUT range instead of repeatedly clicking Append Scan.

Explicit IN always takes priority, including an IN left over from the previous batch. Clear or change it when you want automatic continuation. A set OUT can also prevent continuation beyond that point; clear or advance it for the next block.

Process a video in blocks

1. Set IN and OUT for the first block and run Append Scan.

2. Review the results and optionally export an interim SRT.

3. Set IN and OUT for the next block and run Append Scan again.

4. For a difficult scene, adjust ROI or filters and Append Scan that scene's interval again.

5. Export the complete list when finished.

Caption boundaries crossing a replacement interval are trimmed to preserve the portions outside it. Adjacent matching captions can remain separate entries across different batches; inspect timing at block boundaries.

This workflow continues within the current app session. SubVision does not currently save or reopen a project with scan coverage, and cannot import an SRT to restore an editing session. Exported SRT preserves your deliverable, but does not let you resume the same project after quitting.

Stop and recover

Stop keeps available partial results. When the last caption touches the stopping boundary, it receives an incomplete-scan warning. Append Scan can continue from the saved point; that point may restart at the beginning of the last detected caption to obtain a reliable OUT.

On an early reader failure, the app reports the error and retains the continuation point. With Append Scan, sections outside the processed range are retained. A failure before any frames have been analyzed does not replace the previous list. A scan that has processed frames may already have produced replacement results, so always review a failed or interrupted operation.

7. Set and clear IN and OUT

Scan IN and OUT determine the time interval analyzed by OCR. They are different from Subtitle In and Subtitle Out, which edit the timing of a selected caption.

Set the scan range

Move the playhead to the desired start and click the scan-IN icon immediately before the frame-navigation controls in the bottom bar. Move to the end and click the scan-OUT icon after those controls. The shortcuts are Command-[and Command-]; the physical bracket-key positions are also handled for supported keyboard layouts.

Status shows the set timecodes in the app's highlight color. IN must be earlier than OUT, and the effective range must be within the loaded video. With no IN, Scan starts at zero; Append Scan follows its continuation rules. With no OUT, either operation can run to the end of the video.

Clear one point or both

- Click the small x beside IN in Status to clear only IN.
- Click the small x beside OUT to clear only OUT.
- Click Clear Range to remove both scan points.
- Right-click either scan-point button in the bottom bar for Clear IN or Clear OUT.

Clearing OUT restores the video end as the range boundary. Clearing IN restores the operation's default starting behavior; for Append Scan this may be a continuation point rather than zero. Range controls are disabled during scanning.

8. Review, search and warnings

Selecting a caption seeks the video to its start. Review the words against playback, including captions with no warning: a high confidence score is not proof that the text is correct.

Search captions

Search filters caption text without changing the stored list. Matching is case-insensitive and accent-insensitive. The count shows the visible captions relative to the full list. Use the x in Search, or Escape while editing the search field, to clear the query.

When a filter hides the selected caption, its selection is cleared. Search resets when another video is opened. Export SRT always exports the full list; Search is a review tool, not an export-range selection.

Warnings only

Warnings only is visible when at least one caption has a warning. Enable it to show captions with confidence below 95%, failed recognition near their time range, or an incomplete scan at the stopping boundary. It combines with Search, so both conditions must match. If the last warning disappears, the control disappears and the filter turns off automatically.

The orange warning icon's tooltip explains the reason. A percentage identifies low confidence; failed or incomplete scan warnings are separate reasons. A stopped scan does not automatically mark every completed caption. If recognition fails without producing a caption, the issue appears in the scan report and cannot appear as a caption in the filtered list.

Scan-warning metadata is recorded by current scans. Changing caption text manually does not certify a scan warning as resolved. A successful re-read or replacement scan replaces obsolete scan-warning metadata in the processed range, although the new caption can still have low confidence.

9. Edit captions and re-read

Correct text

Hover a caption and click its edit icon, or use Return when the list has keyboard focus, to start editing. Type the corrected text and confirm with Return. Escape cancels editing. Leaving the text field also commits a valid edit. Delete is available from the caption's context menu.

Clear All removes all caption results from the current list. Export first if you need to keep them. There is no project-level undo or restore workflow for a cleared list.

Adjust caption timing

Select a caption and move the playhead to the required start or end. Subtitle In (I) changes its start; Subtitle Out (O) changes its end. These controls are in the caption sidebar. If no caption is selected, they can create a manual caption at the playhead; add its text and give it a valid end time before export.

The left/right arrows step one frame when no text field is being edited. The bottom-bar circular controls also offer frame steps and one-second jumps. Keyboard shortcuts should not be used to change timing while typing caption or Search text.

SubVision trims conflicting subtitle boundaries to avoid overlapping entries and removes invalid zero-duration entries where sanitization is applied. Check adjacent captions after changing timing. Export Offset is applied later during export, so exported timecodes can differ from the displayed timecodes.

Re-read Selection

Select a caption and choose Re-read Selection, or right-click it and choose Re-read. This scans a local interval around its start and end, with a small margin, using current settings. Results replace caption portions in that local interval; nearby portions inside it can also be affected.

A stopped or failed re-read preserves the existing captions and marks the selected caption with the relevant scan warning. A successful re-read may update the text, leave it unchanged, or remove it when the interval contains no confirmed text. The app highlights an updated caption and reports the outcome. Review the new selection and boundaries rather than assuming a re-read must preserve one caption with the same text.

10. Export SRT and leave safely

Export SRT writes a UTF-8 SubRip file with numbered entries and millisecond timestamps. It includes every valid, nonempty caption in the list, even when Search or Warnings only hides it. Entries are ordered by start time, invalid/empty entries are excluded, and overlaps are sanitized after applying Export Offset.

Timing offset

Export Offset compensates for a consistent timing error across the list. Its standard value is -0.30 seconds, shifting exported cues earlier. The displayed list remains unchanged. Negative start times are clamped to zero. Compare an exported test in your destination editor or player before accepting a long extraction.

Changing text, timing or an export-affecting setting after saving makes the current export out of date. Save again to keep the SRT aligned with the current result. Scan-warning metadata itself is not written into SRT.

Quit or close the window

If captions need export, closing the main window or quitting asks you to export, discard and quit, or cancel. Export opens the SRT save panel after the confirmation dialog has closed. The app exits only after a successful export with no newer pending caption changes. Canceling the save or a write failure leaves the app open.

Stop a running scan before opening another video or quitting. Leaving the app discards its in-memory session even when the SRT has been exported. There is no saved project file in this version; exporting is essential when your captions need to survive the session.

11. App Store and Privacy

Local processing and file access

SubVision recognizes subtitle text locally on your Mac using Apple's Vision framework. The OCR workflow does not upload source video or extracted captions to external servers. Opening the online Terms & Privacy page launches your browser; consulting the bundled Help does not require an internet connection.

The app runs in the macOS sandbox. Access to source videos and exported SRT destinations follows files you select through Open Video, drag and drop, or the save panel. Keep source files and connected drives available while scanning. Export SRT writes to the destination you choose; the app does not publish or share that file automatically.

Mac App Store purchases

For a copy purchased from the Mac App Store, purchases and account licensing are managed by Apple. SubVision does not request your Apple Account password or process payment details. Use the App Store purchase history to download the app again on eligible Macs signed in to your account, subject to Apple's availability and account rules.

Export captions before changing video or leaving the app. An App Store purchase or reinstallation does not restore an unsaved in-memory editing session.

Terms and Privacy

Read the developer's Terms of Service and Privacy Policy at the following address. The page covers the website and the developer's apps; product-specific disclosures should be read for the app they name. It also explains how to contact the developer about privacy requests.

https://www.marcobucci.eu/terms-privacy.html

The local-processing description above applies to SubVision's OCR workflow. Disclosures about other products' AI or translation services do not describe a feature of SubVision.

12. Troubleshooting and diagnostics

No captions found

Check that the ROI includes actual subtitle pixels on several sampled frames. Lower Min Confidence or Confirm Frames for a short test, and decrease the sampling interval for brief captions. Choose a supported language model and inspect the crop after changing filters. An interval without subtitles can correctly produce zero captions.

Random characters or missed words

Exclude logos and background text with a tighter ROI. Test a suitable language group and primary language. Increase confirmation for flickering detections, or lower it for very brief captions. Try contrast, grayscale or threshold on the crop. For proper names or unusual spelling, compare Language correction on and off.

Append Scan starts at an unexpected time

Check Status for a lingering IN or OUT. An explicit IN wins over continuation, and OUT limits the end. Clear the relevant point or set the next block explicitly. With no IN, interrupted scans can repeat the final caption to repair its ending; completed scans continue from recorded coverage, including text-free intervals.

Playback shortcuts do not respond

Click the video player to restore its keyboard focus before using playback shortcuts such as Space. When Search or a caption text field is active, typing should remain in that field. Use the explicit bottom-bar navigation controls when focus is uncertain.

Inspect diagnostics

Open Settings and click the stethoscope icon. The centered Diagnostics Report includes source-file access, bookmark and media status together with session diagnostics. Use Copy to capture it, and Done or Escape to close the panel. Errors and discarded low-confidence frames are reported separately; no-text frames alone are not a system failure.

For support, visit marcobucci.eu and include the app version/build, manual revision, macOS version and a description of the operation. Do not send private source footage unless you choose to share it. SubVision is an independent application and is not affiliated with or endorsed by Apple.

13. Keyboard shortcuts and Help

Shortcut	Action
Command-O	Open Video, with protection for unexported captions.
Command-comma	Open or close Settings.
Command-E	Export the complete caption list to SRT.
Command-period	Stop the current scan.
Command-[/ Command-]	Set scan IN / OUT at the playhead.
Left / Right arrow	Step backward / forward one frame outside text editing.
I / O	Set Subtitle In / Out or create a manual entry when none is selected.
Return	Start or confirm text editing when the caption list/editor has focus.
Escape	Cancel text editing, clear Search when focused, or close supported panels.
Command-Q	Quit, checking for captions that need export.
Command-?	Open SubVision Help.

Playback shortcuts depend on the native player's focus. Single-key timing/navigation shortcuts are suppressed while editing a text field and during scanning. Scan IN/OUT and caption In/Out are separate operations.

Consult Help

Choose Help > SubVision Help for the offline manual. Search matches chapter and section content; multiple words narrow the results to topics containing all of them. Contents returns to the chapter list, and Previous/Next navigates between chapters. Selecting a section result takes you to that section within its chapter.